INFLUENCES ON EXPERT INTELLIGIBILITY JUDGMENTS OF SCHOOL-AGE CHILDREN'S SPEECH

by
JILL R. POTRATZ

A DISSERTATION
Presented to the Department of Linguistics
and the Division of Graduate Studies of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

December 2022

DISSERTATION APPROVAL PAGE

Student: Jill R. Potratz

Title: Influences on Expert Intelligibility Judgments of School-Age Children's Speech

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Linguistics by:

Melissa A. Redford	Chairperson/Advisor
Gabriela Pérez Báez	Core Member
Melissa Baese-Berk	Core Member
Stephanie DeAnda	Institutional Representative

and

Krista Chronister	Vice Provost for Graduate Studies

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded December 2022

DISSERTATION ABSTRACT

Jill R. Potratz

Doctor of Philosophy

Department of Linguistics

December 2022

Title: Influences on Expert Intelligibility Judgments of School-Age Children's Speech

Speech-language pathologists (SLPs) make impressionistic intelligibility judgments as part of an evaluation of children for speech sound disorder. Despite the lack of formalization, it is an important measure of choice for SLPs, going beyond single-word standardized measures by using spontaneous speech to assess functional communication. However, spontaneous speech introduces sources of error and bias in the listener. This dissertation argues that impressionistic intelligibility judgments are influenced by listener-dependent factors due to their subjectivity. To identify potential sources of error and bias, speech data were collected from four school-aged child groups: typically developing monolingual, children with speech sound disorder, typically developing Spanish-English bilingual (i.e., an accent familiar to the study's listeners), and typically developing Mam-English bilingual (i.e., an accent unfamiliar to the study's listeners), in two school-age groups. Perceiver data were collected from two listener groups (i.e., expert [SLP] and lay). Listeners provided baseline measurements of lab-based intelligibility scores and comprehensibility ratings by orthographically transcribing and rating audio recordings of experimentally controlled utterances. Listeners also made impressionistic global intelligibility assessments after viewing video recordings of children's spontaneous speech. Findings showed differences between expert's and lay listener's global intelligibility assessments; however, experts were no better than lay listeners at discerning between age and speaker groups. Of the four speaker groups, there was a significant effect of the Mam-English bilingual speaker group on global intelligibility assessments. Relationships were found between global intelligibility assessments and both the lab-based intelligibility measure and the comprehensibility rating, indicating impressionistic judgments tap into both speech signal features and the understandability of speech. Surprisingly, the age and linguistic ability of the child speakers were not significant factors on global intelligibility assessments, so perhaps listeners were making accommodations for these differences in their assessments. These findings indicate the need for increased training of SLPs to reduce error and bias in their speech intelligibility judgments, as well as the need for further research to improve its objectivity.

CURRICULUM VITAE

NAME OF AUTHOR: Jill R. Potratz

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
University of Wisconsin-Madison
Marquette University, Milwaukee, Wisconsin

DEGREES AWARDED:

Doctor of Philosophy, Linguistics, 2022, University of Oregon
Master of Science, Communication Sciences and Disorders, 1992, University of Wisconsin-Madison
Bachelor of Science, Speech Pathology and Audiology, German minor, 1988, Marquette University

AREAS OF SPECIAL INTEREST:

Assessment of Child Speech and Language
Bilingualism in Children

PROFESSIONAL EXPERIENCE:

Communication Sciences and Disorders Assistant Professor, Augustana College, August 2022-present

Communication Disorders and Sciences Lecturer and Clinical Instructor, University of Oregon, 2015-2017

Communication Disorders and Sciences Lecturer, California State University, East Bay, 2015

Communication Disorders and Sciences Lecturer and Clinical Instructor, San Francisco State University, 2014-2015

Communication Disorders and Sciences Lecturer and Clinical Instructor, San Jose State University, 2013-2015

GRANTS, AWARDS, AND HONORS:

Most Valuable Professor Award, University of Oregon Women's LaCrosse Team, 2016

Employee of the Month, Alameda Unified School District, 2011

Cancer Prevention Summer Assistantship Program, University of Wisconsin Clinical Cancer Center, 1992

PUBLICATIONS:

Potratz, J. R., Gildersleeve-Neumann, C., & Redford, M. A. (2022). Measurement Properties of mean length of utterance in school-age children. *Language, Speech, and Hearing Services in Schools. 53*(4), 1088-1100.

Jasso, J., & Potratz, J. R. (2020). Assessing speech sound disorders in school-age children from diverse language backgrounds: A tutorial with three case studies. *Perspectives of the ASHA Special Interest Groups*, *5*(3), 714–725. https://doi.org/10.1044/2020_PERSP-19-00151

Potratz, J. R., Dengel, G., Robbins, J., & Brooks, B. (1993). Movement analysis of oropharyngeal dysphagia: An interactive computer assisted approach. *Journal of Medical Speech/Language Pathology*, *1*(1), 61–69.

Potratz, J.R., Dengel, G. & Robbins, J.A. (1992). A comparison of swallowing in three subjects using an interactive image processing system. *Proceedings of the Fifth Annual IEEE Symposium on Computer-Based Medical Systems*, 115-122.

ACKNOWLEDGMENTS

I wish to express sincere appreciation to Professor Redford for her assistance in the preparation of this manuscript. In addition, special thanks are due to Dr. Gabriela Pérez Báez, Dr. Nora England, and Ms. Ana Maria Dudley, whose expertise and familiarity with the Mam community was helpful during the preparation and running phases of this undertaking. I also thank Ms. Elda Carrillo for her assistance with participant recruitment and Mam interpretation. Thanks are extended to Dr. Stephanie DeAnda and the Early Dual Language Development Lab for access to a database of Spanish-speaking families. Ms. Julisa Silva-Ramirez is appreciated for her assistance with participant recruitment and Spanish interpretation. The investigation was supported in part by grant R01HD087452 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD) awarded to Dr. Melissa Redford at the University of Oregon and a research award from the University of Oregon Department of Linguistics.

To my partner, Lowell Schneider, for his emotional and financial support during the writing of this dissertation.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

CHAPTER I

INTRODUCTION

A child's eligibility for special education services in the public schools is dependent on the evaluation of many professionals. If a speech and language impairment is suspected, the professional who completes an evaluation is a speech-language pathologist (SLP). Impairments with the development of speech sounds are referred to as speech sound disorders (SSD), which occur in 5% of early school-age children (Law et al., 2000; Shriberg et al., 1999). SSDs often impact a child's ability to communicate in the classroom with their teacher and peers. Additionally, the language and literacy skills of children with speech sound disorders have been shown to be impacted into adolescence (Lewis et al., 2015).

As an example, a school-based SLP receives a referral from a Kindergarten teacher who says they cannot understand what their student is saying to them. The SLP's job is to determine if the child is eligible for speech therapy by assessing their speech skills with a variety of assessment tools including standardized measures and an impressionistic judgment of their intelligibility. The SLP suspects an SSD, but the case history reveals that the child is an English language learner in a monolingual English school environment. Their first language is Vietnamese, a language this SLP is not familiar with. Best practices exist for the inevitable times when there is a mismatch in language between the client and the SLP (e.g., McLeod et al., 2017), however SLPs report feeling unprepared and inadequately trained in these assessments (e.g., Arias & Friberg, 2017). Consequently, misdiagnosis (i.e., both under-identification and over-identification) of a speech and language impairment is a problem in multilingual children

(Artiles et al., 2002; Cycyk et al., 2022; Sullivan, 2011). So, it could turn out that this student has an SSD that goes unidentified because the SLP determines sound differences are due to the child learning English. Another outcome might be that this student is identified as having an SSD, when in fact they are acquiring the sounds of two languages which has impacted their intelligibility.

Our hypothetical student was referred since their English was poorly understood by their teacher; they were perceived to not speak clearly. While SLPs are trained listeners, they are human and may be influenced by factors that are not related to SSD. First, they may be influenced by developmental expectations based on the child's age. Second, they may be influenced by the linguistic abilities demonstrated by the child, knowing that children with more complex syntax and higher-level vocabulary use typically have more advanced speech skills (i.e., more clearly articulated speech sounds). Depending on the amount of exposure to English the child has had, their speech may be influenced by cross-linguistic transfer such that their English production may be influenced by Vietnamese (Fabiano-Smith & Barlow, 2010). If the SLP has never heard Vietnamese-influenced English spoken before, they will not be familiar with this accent. This lack of familiarity with an accent is another potential influence on the SLP's appraisal of a child's intelligibility.

What exactly is intelligibility? How can it be better assessed by SLPs? These are questions addressed in this dissertation. The motivation for this study at the highest level, is about reducing sources of error and bias, such as in the targeted factors of age, linguistic ability, and accent familiarity. This can be done by improving the training that SLPs receive regarding speech intelligibility assessments. The overarching study goal is to better characterize the underlying speech and perception factors that influence SLP's intelligibility judgements to improve their use of these judgments in practice. The current study uses speech sample data collected from four groups of child speakers in two age groups (i.e., younger, and older school-age children in: typically developing monolingual English, speech sound disorder, typically developing Spanish-English bilingual, and typically developing Mam-English bilingual groups). Expert (i.e., SLP) and lay listeners provided impressionistic intelligibility judgments of the children's spontaneous speech. For comparison, listeners also made orthographic transcriptions and comprehensibility ratings of semantically anomalous sentences for baseline measures. Once these baseline measures were established, we focused on any differences between listener groups' impressionistic measures. Then, we focused on any influence created by the speaker's age and linguistic background (e.g., differences between the familiar accent of Spanish-English bilinguals versus an unfamiliar accent of Mam-English bilinguals and disordered speech). Last, we focused on the influence of language factors (e.g., syntactic complexity and lexical diversity) using language analyses of the speaker's narrative samples that were completed with transcribed text.

In this dissertation speech intelligibility assessments are the area of specific interest because the spontaneous, connected speech that they utilize, introduces both positive (i.e., it demonstrates the child's functional communication skills) and negative factors (i.e., we do not know what SLPs are listening to) that influence their judgments. So, while the assessment is a vitally important tool, we do not know what it is really measuring. While aiming to examine what an intelligibility judgment captures, we will determine if SLPs are influenced by targeted factors. Therefore, we will determine the factors that influence judgments, the extent of their influence, how the factors relate to an assessment of SSD, and additionally, what factors might not be relevant to SSD.

Following a review of the relevant literature on the importance of speech intelligibility assessments, we discuss the current practice of SLPs and some of the potential influences on the listener. The remaining chapters present and discuss the results obtained for analyses on intelligibility and comprehensibility assessments and the targeted factors that influence impressionistic judgments. This will be followed by a discussion of the conclusions of the current study and some thoughts on suggestions for future research and the training and clinical implications of our findings.

CHAPTER II

BACKGROUND

A key tool that practicing SLPs use in the evaluation of children for speech sound disorder (SSD) is an impressionistic estimate of intelligibility (Gordon-Brannan & Hodson, 2000) that is based on a spontaneous speech sample. These estimates are important because they capture that which is not captured in single-word standardized measures—the child's ability to functionally communicate during connected speech. However, since impressionistic judgments of intelligibility are established with an auditory-perceptual assessment of spontaneous speech by a clinician, they are susceptible to a variety of sources of error and biasing effects of speaker characteristics (Kent, 1996). Training can be used to overcome these problems, but first we need to understand what SLPs are doing in practice. From there we can identify the potential of sources of error and bias in this practice to minimize their undesired effects.

2.1. The Importance of Intelligibility Assessments

Whether developmental, or due to a disorder, children's speech articulation often results in reduced speech intelligibility. Children with speech sound disorder (SSD) have especially reduced speech intelligibility; that is, intelligibility is below age expectations due to speech sound distortion, substitution, deletion, and/or other types of errors (Hustad, 2012; Lousada et al., 2014; McLeod et al., 2013). Children with moderate to severe SSD may have atypical speech motor control that results in especially variable productions that further impact their intelligibility (Grigos, 2016; Namasivayam et al., 2013). Communication success is compromised because the listener has difficulty understanding the child. In other words, the child's speech is misperceived and so is labeled as reduced in intelligibility.

SLPs calculate intelligibility in school-aged children as part of a broader evaluation to determine initial eligibility for special education, but also to plan treatment, or determine the need for continued speech therapy services. A comprehensive initial evaluation for suspected SSD will include several steps to determine whether errors produced by the child are appropriate for their age and linguistic background (e.g., Bleile, 2002; Fabiano-Smith, 2019; Skahan et al, 2007). After gathering the child's case history (including language history and dialect information) from a parent or guardian, the SLP engages the child in the following tasks: an oral-facial examination, an audiological screening, a criterion referenced test (e.g., Percent Consonants Correct [PCC]; Shriberg & Kwiatkowski, 1982), a single-word test that is often standardized (e.g., Goldman-Fristoe Test of Articulation-3; Goldman & Fristoe, 2015), stimulability testing to determine what sounds the child can produce given a model and short instructions, contextual testing to determine performance across different phonetic contexts, and an assessment of the child's ability to produce multisyllabic words. These evaluations also include an intelligibility assessment based on a spontaneous speech sample.

Kwiatkowski and Shriberg (1992) stated that "conversational speech sampling provides the only valid approach for children with phonological disorders" to "classify a speaker's level of unintelligibility, to describe certain linguistic dimensions of unintelligibility, or to attempt to explain the relevant correlates of unintelligibility." Accordingly, eliciting and analyzing a spontaneous speech sample has become the procedure of choice for determining intelligibility. It is a valid means to assess the phonological ability of children with suspected SSD (Morrison & Shriberg, 1992; Shriberg & Kwiatkowski, 1985). Prior to eliciting a spontaneous speech sample, the clinician determines the length and type of sample to be collected (e.g., play-based, conversational, narrative, expository), matched to the age and abilities of the child. The speech sample needs to be long enough to ensure a representative sample of the child's speech skills; the duration will vary (e.g., 5-20 minutes) depending on the verbosity of the child. Play-based samples are elicited from younger, less skilled children, whereas expository tasks (e.g., favorite game/sport task) may be chosen to elicit more complex language from older children (e.g., Nippold, 2009). Conversational samples are elicited with open-ended questions (e.g., tell me about your favorite movie/video game/book) and have the advantage of utilizing topics that may be of high interest to the child and thus elicit more language. Narrative samples, elicited with picture cards or wordless story books/videos, have the advantage of being more structured than conversation and of potentially containing vocabulary with target sounds. Also, a story retelling task may show off the best of the child's linguistic performance as they are likely to have heard modeled vocabulary and syntax (e.g., Serratrice et al, 2015).

In general, spontaneous speech sampling has the advantages of being quick and multipurpose, making it an efficient tool for school based SLPs who conduct assessments and provide treatment for many students – very high caseloads are the norm for school SLPs (e.g., Katz et al., 2010). Time saving strategies are essential to SLPs who serve in the schools and so many children with varied needs. In addition to being a means to an intelligibility assessment, the spontaneous speech sample can also be used to: 1) assess the child's functional communication skills (i.e., are they able to get their message across), 2) identify speech errors, particularly revealing any differences in phoneme production between single words and connected speech, and 3) evaluate language (e.g., syntactic complexity and lexical diversity), voice, fluency, and prosody.

In fact, leaving out a spontaneous speech sample from an SSD evaluation may be detrimental to the purpose of the assessment. Differences between single word productions, like those on an articulation test, and spontaneous speech have been documented (Andrews & Fey, 1986; Iacono, 1998; Yeh & Liu, 2021), indicating either an increase or a decrease in errors depending on the task. Commonly, children perform in an age-expected manner on word- or sentence-level procedures, but not when they connect words into more complex sequences. For example, a child with disordered speech may score in the average range for their age on a single word articulation test but cannot be understood sufficiently during connected speech due to errors related to sentence length and phonetic complexity (Allison & Hustad, 2014; Ertmer, 2010; Morrison & Shriberg, 1992). In some speech and language disordered children it has been shown that as linguistic complexity increases so too do articulation errors (Bernthal et al., 2022; Panagos & Prelock, 1982; Paul & Shriberg, 1982). Ultimately, word- or sentence-level procedures are not valid for the purposes of an intelligibility assessment since they do not elicit the interaction of language, speech, voice, and prosody needed to determine intelligibility (Kwiatkowski & Shriberg, 1992).

Single-word- and sentence-level measures have their place in the assessment. They provide a means for capturing information about the child's phonetic inventory in a systematic way. They give a representative sample of individual consonants and vowels in a context of known targets, so the clinician knows what the child is attempting to say

(Eisenberg & Hitchcock, 2010; Pollock, 1991). Norms or criterion references are provided with standardized assessments which helps the clinician determine whether the child is performing in an age-expected manner. However, single-word tests are only one piece of the assessment puzzle and do not provide information about the child's ability to communicate effectively. Importantly, what is missing from single-word measures is information gained from connected speech. In other words, intelligibility is not just about individual sounds, thus connected speech is required.

The importance of impressionistic judgments of speech intelligibility that SLPs make has been established, but they are problematic for multiple reasons (Gordon-Brannan & Hodson, 2000; Miller, 2013; Skahan et al., 2007). First, there are currently no well-defined methods or validated instruments to measure speech intelligibility in clinical settings (Bernthal et al., 2022). Second, the intended targets in children's utterances are often not known, making it difficult to identify and assess their speech production (Flipsen, 2006; Kwiatkowski & Shriberg, 1992). Third, when measuring treatment progress impressionistic judgments of children's speech may not be sufficiently sensitive to pick up subtle improvements due to speech therapy (Lousada et al., 2014). Lastly, intelligibility measures are notoriously, and inherently limited by their subjective and impressionistic nature (Lousada et al., 2014; Yoder et al., 2016). This level of subjectivity has led several researchers to suggest more than one listener provide a rating to improve the measure's reliability (Bernthal et al., 2022; Hustad et al., 2015; Yoder et al., 2016).

In sum, when SLPs complete a comprehensive assessment of a school-aged child with SSD, or suspected SSD, making a speech intelligibility judgment from a spontaneous speech sample is a method of choice for indicating functional communication ability that is not captured in a single word assessment. However, these intelligibility assessments are problematic for several reasons related to their use of spontaneous speech, making them susceptible to sources of error and the biasing effects of speaker characteristics. To better understand how error and bias are issues with this measure, in the next section we turn to the current practice of speech intelligibility assessments.

2.2. Speech Intelligibility Assessments in Practice

As discussed above, while vitally important, intelligibility assessments are not formalized and so are determined in a variety of ways: 1) impressionistic judgments expressed as either approximations of the percentage of speech understood or values on a rating scale, 2) a quantifiable method like word or sentence identification tasks, or 3) a combination of these methods (Bleile, 2002; Ertmer, 2010; Lousada et al., 2014; Pascoe et al., 2006). Impressionistic judgments are made while listening to live or recorded spontaneous speech and a perceptual judgment is made about the amount (e.g., percentage) of speech understood (Bernthal et al., 2022; Bleile, 2002; Shipley & McAfee, 2009). Word and sentence identification tasks are more time consuming as they require recording the child and subsequently having an additional listener attempt to identify the words or utterances.

To document the current use of intelligibility assessments in practice, SLP participants in the current study (*n* = 43) as well as additional school-based clinicians from the San Francisco Bay Area (via email prior to the current study; *n* = 11) were asked about how they measure and define intelligibility in their child clients. A sample of responses are presented in Table 1.

Table 1. A sample of survey responses

How do you measure intelligibility? (*N* = 54)	How do you define intelligibility? (*N* = 54)
Global approximations (*n* = 24):	Reference comprehensibility (*n* = 35):
"*I would get an audio recording of a language sample consisting of 100 words and ask an unfamiliar person (an slp friend) what they understood, then calculate a percentage. And do the same with a child's caregiver to compare with a familiar person. Honestly I don't always have time to do this but if I had to formally assess, this is what I would do. I would get permission from parent to obtain and share the recording and make sure no personal information was shared. If this was not a big area of concern, I would interview familiar people (teachers, aides, ect...) and ask them to "ball park" how much of their speech is understood with an estimated percentage.*"	"*how much of a person's spontaneous speech I understand*"
"…If you are asking more about how I determine % intelligibility - it's again admittedly more subjective and a gut feeling than anything objective"	*"my ability to comprehend their spoken utterances and their ability to effectively express wants, needs and direct the actions of others using verbal speech"*
Quantified measure/percentage; (*n* = 17):	*"the ability to be clearly understood while speaking, 80-90% of the time, by familiar and unfamiliar listeners"*
"*sometimes, I will record a student both doing a picture description and in conversation and transcribe it and measure how many utterances I understood vs how many contained unintelligible words to get a percentage*"	*"ability for a listener to understand what someone is saying"*
"I usually recorded samples across 2-3 trials and used +/- to mark each word then divided by 100 to get a percentage of "intelligible/unintelligible" speech…"	*"a measurement of how comprehensible someone's speech is"*
"I use PCC to report objective intelligibility, and subjective ratings of poor/fair/good with familiar and unfamiliar listeners"	*"% intelligible (estimate of the % of words understood) under different conditions...familiar listeners in known context, unfamiliar listeners in unknown context, single words, connected speech during formal language activities and during play. etc. I'll record a language sample. I try to get an idea about intelligibility at my first or second session...before I would be considered a familiar listener"*
Rating scale; (*n* = 11):	*"how well the child can be understood by unfamiliar listeners when the context is not known. I guess I define normal speech intelligibility as 90% intelligible by age 3"*
"*I gather informal ratings too from parents (how well do you understand your child on a scale of 1-10, how well do strangers understand your child on a scale of 1-10)"*	*"the proportion of comprehensible speech to the total of the language sample"*
	"how accurately I perceived the attempted spoken message' (if I quantify it with 'with a trained, unfamiliar listener' for an initial, or familiar if a tri). I admit it's not a very defined concept - just a general 'how often do listeners understand the child's attempts to speak'"
	Reference to percent correct (*n* = 5): "*Percentage correct*"

Table 1 continued

How do you measure intelligibility? *(N* = 54)	How do you define intelligibility? (*N* = 54)
"Subjective ratings of poor/fair/good with familiar and unfamiliar listeners"	Reference to clarity of speech (*n* = 2):
	"how clearly sentences are produced"
Norm-referenced assessment; (*n* = 8):	Reference to articulation and prosody (*n* =1):
"I used % intelligibility and usually an artic test like the GFTA"	*"Intelligibility involves articulation, prosody, voice, and semantics and syntax. Pitch, rate, volume and speech sound errors affect one's intelligibility, along with word and sentence errors"*
"For assessment, I use a combination of standardized assessment and interview. For example, the CAAP or GFTA-2 + asking teachers/parents about how often they feel like they understand what the child is saying. Sometimes I will also make an informal speech transcription (not using IPA, just phonetic spelling) as a visual for how often and what phonemes are produced differently."	
"*I use SALT at times…*"	

The survey results confirm there is no single method for measuring speech intelligibility and that most clinicians use some version of a subjective estimate. More than half of the SLPs reported measuring intelligibility impressionistically, with a global approximation or rating scale (see left column of Table 1). Several reported using a quantitative method with a norm-referenced assessment, or by calculating the percentage of words understood from either transcribed speech or a single-word assessment. For example, several surveyed mentioned using SALT (i.e., Systematic Analysis of Language Transcripts, Miller & Iglesias, 2019) which provides the percent of intelligible utterances and the percent of intelligible words as part of its standard measures report. Several SLPs reported using a combination of these methods. Turning to the definition of intelligibility (see right column of Table 1), most of those surveyed indicated that they defined intelligibility in terms of their understanding of speech. Some SLPs indicated the importance of clarity of speech and "percent correct." Thus, SLPs define intelligibility as more than simply the quantity of speech errors; they, instead put a focus on functional communication, or the comprehensibility of spontaneous speech. A similar definition of intelligibility can be found in clinically-focused textbooks: “the amount of speech a listener understands” (e.g., Bleile, 2020). By contrast, the research literature defines intelligibility as the extent to which an acoustic signal, generated by a speaker, can be correctly recovered by a listener (Hustad, 2008; Hustad & Cahill, 2003; Hustad et al., 2015; Ishikawa et al, 2021; Kent et al., 1989; Mahr et al., 2020). In other words, the research literature puts an emphasis on the speech signal, which can be objectively measured using speech acoustics; practicing SLPs focus on comprehensibility, which is estimated based on their listening to a speech sample.

These survey results, along with the procedures described in the previous section, and the literature (e.g., Pascoe et al, 2006), demonstrate that SLPs typically measure and define speech intelligibility with an emphasis on the listener. To that point, the definition of comprehensibility found in the literature aligns with how SLPs define intelligibility in practice: the listener’s perception of how easily they understand the speaker’s message (Derwing & Munro, 1997). Thus, the “intelligibility” in impressionistic intelligibility assessments is really “comprehensibility.” While this can be construed positively since it is about the functional communication of the child, it also opens the door to many problems, not least of which is the fact that the assessment is not based solely on speech; it is also likely influenced by language factors. Also, clinicians may bring their own expectations, or listener biases, to the comprehension task. Before examining these potential biases, we will contrast the ideas of speaker-focused intelligibility with listener-focused comprehensibility.

2.3. Intelligibility versus Comprehensibility

An intelligibility assessment completed with a speaker-focused lens is especially relevant for a diagnosis of SSD since a child must ultimately approximate the adult model to effectively communicate with the adult. The focus is therefore on variables like articulatory precision (e.g., presence of distortions, substitutions, additions, or omissions) and prosodic features (e.g., speech rate, intonation, vocal volume). For example, consonant clusters and/or monophthongs in closed syllables are specific phonetic features that have been shown to negatively impact intelligibility in young (i.e., 4–5-year-old) children (Weston & Shriberg, 1992). Suprasegmental speaker variables also impact intelligibility, including speech rate and vocal intensity, as one SLP pointed out in the survey (See Table 1, right column, last entry). Listener ratings of second language speech support this observation: second language speech that is especially fast or especially slow results in decreased listener ratings (Munro & Derwing, 2001). Likewise, vocal intensity impacts the speech intelligibility of children with dysarthria such that greater intensity increases speech intelligibility (Levy et al., 2017).

Speaker-focused intelligibility is illustrated in the left triangle of Figure 1: it is defined as the signal-dependent, bottom-up factors that affect listener's perception, including articulatory precision and prosodic features. Comprehensibility on the other hand (see Figure 1, right triangle) is defined as knowledge-dependent, top-down factors that are due to the listener and influence how the listener hears speech. This includes the listener's existing knowledge (e.g., accent familiarity) and any known or unknown biases based on speaker characteristics (e.g., knowledge of client's history and/or the speaker's physical appearance).

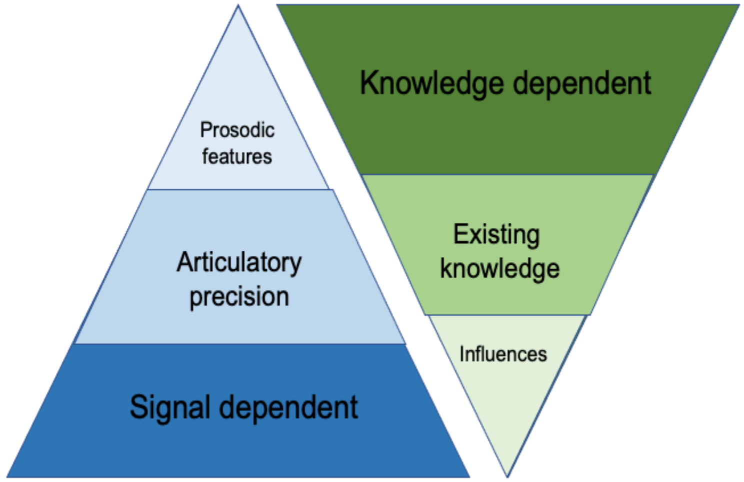


Figure 1. Types of listener processing: Intelligibility (left triangle) is a bottom-up approach derived from speaker variables; comprehensibility (right triangle) is a top-down approach influenced by listener variables.

How might top-down influences on the listener, that are potentially irrelevant to SSD, be separated from information that is relevant? One way is to use item identification measures that can control for things like word familiarity (e.g., Bent and Bradlow, 2003) and/or developmental appropriateness, such as the lab-based intelligibility scores used in this dissertation (LIS; see Chapter 3). The procedure for an item identification measure controls for language factors because it involves recording the speaker saying pre-determined words or sentences. Listeners then orthographically transcribe what they hear based on the recordings, which controls for visually-based factors. Scores are then derived based on the number of key words correctly identified. But, since this measure is perceiver-based it is still impacted by some listener properties, such as a lack of familiarity with nonnative English. Nonetheless, it is a more controlled measurement than derived from rating intelligibility based on an in-person elicitation of spontaneous speech.

Item identification has been used to estimate intelligibility of child talkers from various populations. Hustad and colleagues (2015) used the measure to investigate speech intelligibility scores within and between listeners who orthographically transcribed 1-4 words of imitated, developmentally appropriate (i.e., lexically, phonetically, syntactically, and morphologically), speech generated by either typically developing or speech motor impaired 5-year-olds. They found these intelligibility scores to differentiate between the typically developing and disordered speaker groups even though there was considerable variability within and between listeners.

These findings indicate the validity of item identification tasks since they remove some top-down influences on the listener, making them more objective compared to rating scales (Kent et al.,1989; Kwiatkowski & Shriberg, 1992). While some argue that item identification tasks are time-consuming and thus not practical from the point of view of a busy school-based clinician (e.g., Bleile, 2002), others might argue that the time cost may be worth having better outcomes and services. But another way to mitigate top-down influences of the listener on intelligibility judgments is to identify these influences and improve training to decrease them.

If an intelligibility assessment has a listener-focused lens (i.e., comprehensibility; see Figure 1 right triangle), it is inherently influenced by the listener's existing knowledge, including things like the speaker's age, language ability, and accent (Adank et al., 2009; Allison & Hustad, 2014; Derwing & Munro, 1997; Flipsen, 2006; Gass & Varonis, 1984; Hustad et al., 2012; Munson et al., 2010; Weston and Shriberg; 1992), each of which are included in the current study's approach. For example, as children's speech develops with age, they become more intelligible (e.g., Flipsen, 2006); accordingly, adults rate children's speech as more accurate when they believe the children are older (Munson et al., 2010). This indicates that a listener's expectations, based on prior knowledge may influence the degree to which they rate children's speech as comprehensible. Contributing to these expectations is the level of experience and exposure that the listener has to a particular variable, including experience-based familiarity with a particular speaker or accent. This type of prior knowledge has an impact on speech perception because we integrate our existing knowledge with incoming sensory information (e.g., Sohoglu et al., 2012). As a result, a listener who is more familiar with a particular person's speech will understand it better; a child's caregiver is much better at recognizing words spoken by their child compared to a clinician or less familiar listener (Baudonck et al., 2009; Goehl & Martin, 1987; Tjaden & Liss, 1995). Similarly, an individual listener's intelligibility ratings improve following repeated exposure to specific speakers (e.g., Hustad & Cahill, 2003).

The listener-focused nature of intelligibility judgments means that experience likely introduces a certain amount of bias – both positive and negative – to these assessments. Given this likelihood, one of the research questions addressed in this dissertation is: Does the training that an SLP receives, and their more extensive experience with children's speech, effect how they make global intelligibility assessments compared to lay listeners? It is hypothesized that experience improves SLP performance and so that the inexperienced listeners will produce less accurate and reliable results than SLPs, who are expert listeners. The next section discusses what biases the listener might bring to intelligibility assessments in addition to that which is strictly relevant to SSD.

2.4. What Listener Expectations Bring to Intelligibility Assessments

2.4.1. Expectations Based on the Child's Age

Children's speech becomes more intelligible to adult listeners with age (e.g., Flipsen, 2006; Hustad et al., 2020). Typically developing older children, who possess more diverse and complex grammaticality and vocabulary, are perceived as more intelligible compared to younger children. Certainly, their articulation becomes more stable and less variable across utterances (Allison & Hustad, 2014; Hustad et al., 2020). As a result of known age-related changes in speech articulation, adults rate children's speech as more accurate when they believe the children are older (Munson et al., 2010). The two age groups of children in the current study (younger group with a mean age of 6;1 and an older group with a mean age of 7;8) are distinct by the fact that the older group is well beyond the developmental age when they are expected to be 100% intelligible, whereas the younger group is not (Coplan & Gleason, 1988; Flipsen, 2006). Thus, in the current study, we anticipate age-related differences in the global intelligibility judgments.

2.4.2. Expectations Based on the Child's Linguistic Ability

This subsection begins with a description of language assessment procedures that SLPs use to characterize the receptive and expressive language skills of their child clients. This is followed by a discussion of how linguistic complexity might impact the listener due to their expectations of how a child's speech may sound.

Comorbidity of SSD and language delay is high (40.8% at 4-years-old, Eadie et al., 2015; Shriberg et al., 1999), so SLPs often complete an assessment of receptive and expressive language skills with children who have suspected SSD. A comprehensive language evaluation often includes a standardized assessment tool (e.g., Clinical Evaluation of Language Fundamentals [CELF-5], Wiig et al., 2013; Comprehensive Assessment of Spoken Language-Second Edition [CASL-2], Carrow-Woolfolk, 2017) along with additional tests like a vocabulary assessment (e.g., Receptive and Expressive One-Word Picture Vocabulary Tests- Fourth Edition, [EOWPVT-4, ROWPVT-4], Martin & Brownell, 2011) or the elicitation of a language sample for analysis, similar to the spontaneous speech sample described above. Analysis of a language sample (e.g., with software such as the Systematic Analysis of Language Transcripts [SALT], Miller & Iglesias, 2019, that automatically calculates several measures) provides a way of accessing information about a child's language development that does not rely on a standardized measure. Language assessments of multilingual children, which occur in all of the child's languages, may involve the same tools listed above but, as with standardized articulation tests, scores are not valid since most tests are normed on monolingual populations.

The interpretation of this type of language evaluation is coordinated with an analysis of the child's speech development (i.e., articulation and phonology) with the knowledge that there may be complex interactions between speech and language abilities that effect speech intelligibility (Hustad et al., 2012). It is well established that strong language skills enable strong speech skills because of more speech practice. In other words, stronger language abilities and speech intelligibility are related. Children with larger vocabularies are better at nonword repetition tasks (Edwards et al., 2004; Gathercole et al, 1999), providing further evidence for the connection between strong speech skills and strong language skills. Additionally, there is some evidence that language complexity (i.e., more complex sentences; longer utterances), has a negative impact on speech intelligibility since complex language entails more complex phonology (Allison & Hustad, 2014; Hustad et al., 2012; Monsen, 1983; Weston and Shriberg, 1992). This research is primarily with disordered populations, however.

In multilingual populations, linguistic complexity may have a different impact on speech intelligibility. Nip and Blumenfeld (2015) examined simple and syntactically complex sentences in both English and Spanish produced by native English adult speakers who were learning Spanish. They found that syntactic complexity affected speech motor control in their participants, suggesting that second-language production requires greater cognitive resources than native language production. Given this information we might presume that speech intelligibility will be negatively impacted as complexity increases in second language talkers. However, since stronger language skills are most associated with stronger speech skills overall, the results might go the other way.

To summarize, linguistic factors, such as the syntactic complexity and lexical diversity used by the child speaker, influence the listener such that as linguistic complexity increases, so too do expectations for intelligibility. Thus, it is predicted that listeners will be influenced by more complex language and expect higher intelligibility. This leads to the current study question that asks about the impact of linguistic complexity on global intelligibility assessments.

2.4.3. The Impact of the Child's Accent

The impact of child's accent on the listener's intelligibility assessment is correlated with the extent to which the listener is familiar with the accent in question.

Unfortunately, clients are frequently matched to SLPs will little exposure to accented language, resulting in a client—SLP mismatch. The problem of mismatch is severe because of the workforce diversity problem in the field of speech-language pathology. To understand the impact a child's accent may have on the assessing SLP, bilingual assessment procedures are reviewed. This is followed by a discussion of how accent is defined and a description of more general impacts on the listener due to their expectations.

The key difference between a speech sound disorder assessment for monolingual children (see section 2.1.) and bi- or multilingual children is that assessments for multilingual children are conducted in all their languages. While this can be challenging and time consuming, these assessments are within an SLP's scope of practice. The evaluation will include a multilingualism questionnaire to gather information about the child's language history from the parents/guardians. The language history includes when the child was first exposed to their languages and the current levels of input (e.g., what the child hears) and output (e.g., what the child produces) of each language. Standardized assessment tools may be used to assess articulation skills but if they are normed on monolingual populations, scores are not reported (e.g., Hasson et al., 2013). Monolingual speech assessments in languages other than English exist (McLeod, 2012a), but the list of these is far from exhaustive. The assessing SLP may even have to create their own sampling tool in the child's native language if one is not already available (McLeod, 2012b). This was the case in the current study where an articulation screening tool was created for the Mam-English bilingual children.

When determining whether a multilingual child has a speech sound disorder, the goal is to discern whether sound differences are a result of the child acquiring more than one language or due to a speech impairment (e.g., Fabiano-Smith et al., 2021; Jasso & Potratz, 2020; Preston & Seki, 2011). To achieve this goal, the evaluation is conducted in all the child's languages, often with the assistance of an interpreter if there is a mismatch between client and SLP, and the SLP must interpret the results based on the phonetics and phonology of the child's native language. They must determine whether sound differences are due to a disorder or is what would be expected given their specific linguistic community (e.g., Jasso & Potratz, 2020).

The sound differences referred to here may be labeled as the child's accent. There is disagreement in the literature on how to define accent (Ockey & French, 2014), but for the purposes of this dissertation it is defined in terms of speaker differences and listener's perceptions, following Derwing and Munro (2009): "the ways in which their speech differs from that local variety of English and the impact of that difference on speakers and listeners." The impact of a speaker's accent on the listener has been widely studied in recent years. For instance, it has been found that listener performance on word identification tasks decreases with accented speech and the inclusion of multiple accents (i.e., versus a single accent condition) further impedes listener performance (Bent & Frush Holt, 2013). Further, acoustic mismatches between the listener's expectations about what one's speech will sound like, and the actual production of foreign-accented speech leads to a reduction in the intelligibility of the speech. Indeed, understanding accented speech requires greater listening effort and places a greater cognitive load on the listener (Brown et al., 2020; McLaughlin & Van Engen, 2020; Van Engen & Peelle, 2014).

In contrast, familiarity of accent is correlated with understanding such that the more familiar a listener is with an accent, the more comprehensible the accented speech will be (Adank et al., 2009; Derwing & Munro, 1997; Gass & Varonis, 1984). For example, Huang and colleagues (2016) found accent familiarity to facilitate the identification of a particular accent (i.e., Spanish, or Chinese) and that the perceiver's speaker ratings of "overall English proficiency" and "foreign accents" were more lenient toward speakers with familiar accents. So, familiarity with a speaker's accent is a listener-focused influence that facilitates understanding.

What other factors facilitate or inhibit a listener's understanding of accented speech? There may be biasing effects of speaker characteristics that occur when listeners hear children's speech. While comprehensibility generally improves when the listener can view the speaker's face, to use lip reading or other facial cues to enhance their understanding (Brown & Strand, 2019; Hustad, 2006; Ma et al., 2009; MacLeod & Summerfield, 1987; Monsen, 1983; Sumby & Pollack, 1954), when trained and untrained listeners were provided with visual information (i.e., physical appearance) about a child speaker's race, it impacted their assessment of speech production accuracy (Evans et al., 2018). So if a listener expects a child's speech to sound a particular way based on their physical appearance, they anticipate and make judgments based on those expectations. In particular, listeners' beliefs about a speaker's ethnic and/or linguistic background influence the way they perceive speech by either inhibiting or facilitating accurate perception (Kutlu et al., 2022; Melguy & Johnson, 2021; Vaughn, 2019).

Listeners who are not practiced in hearing nonnative English perform poorly in tasks that ask them to understand nonnative English (e.g., Derwing & Munro, 1997; Munro & Derwing, 1995). For example, monolingual listeners judge nonnative speech to be less intelligible than bilinguals do (Fuse et al., 2018). This is particularly important in the SLP population of the United States that is primarily monolingual; only 8.2% of American Speech-Language-Hearing Association's [ASHA] members are multilingual service providers (ASHA, 2021). So, the majority, monolingual SLPs may be judging the intelligibility of nonnative speech more harshly than warranted. In addition to a general prejudice towards nonnative speakers of English (Hansen & Dovidio, 2016), a strong implicit bias against immigrants has been shown in the majority of school-based SLPs surveyed by Nelson and Wilson (2021). This bias impacted the prioritization and utilization of best practices when assessing mulitlingual children. Thus, improving the training of SLPs to conduct evaluations of bilingual children, one of the goals of this dissertation, is crucial since by virtue of their demongraphics, SLPs are especially likely to lack adequate experience with nonnative English speakers.

Contributing to these biases is a lack of resources required for bilingual assessments (i.e., specialized speech assessment tools, phonological inventories, and developmental norms), particularly those being completed with children who speak lesser-known languages. The extant literature on SSD in multilingual populations is focused largely on Spanish-English bilingual children (e.g., Fabiano-Smith & Goldstein, 2010; Gildersleeve-Neumann et al., 2008; Goldstein & Bunta, 2012; Goldstein et al., 2005). So, while there is ample information on Spanish-English bilingual children, there is little, or sometimes no information available to SLPs about children who speak other languages. For example, the ASHA website that lists phonemic inventories and cultural and linguistic information (American Speech-Language-Hearing Association, n.d.) contains information about only 18 languages even though there are approximately 430 languages spoken or signed in the population (This lack of resources holds true for the Mesoamerican language, Mam, which is included as a focus of the current study.

Mam is spoken in Guatemala and parts of southern Mexico and is one of the largest of the Mayan languages with close to a half million speakers (Aissen et al., 2017). The Guatemalan migration to and settlement in the Pacific Northwest of Mam-speaking people began with the civil war in 1980 and continues to the present (Stephen, 2017). Stephen reports that many of the women and children who arrived in Oregon between 2013-2016 were linked to male family members who were working in agriculture and forestry since the 1990s or early 2000s and more recently (2004-2017) women have been fleeing Guatemala to escape violence due to drug, gang, and paramilitary violence. This is not unique to the Mam population, there are large social groups across the United States that are disadvantaged by displacement and present something new to the monolingual English communities where they reside (e.g., Somalis in Minnesota, Darboe, 2003).

One issue that this has engendered is that indigenous people from Mesoamerica, including Mam speakers, living in the United States, are frequently assumed to speak Spanish (K. Mitchell, personal communication, February 4, 2022) and are often treated homogeneously with generic labels, like “Mexican” or “Latinx.” Important to the current study, Mam is an indigenous, Mayan language, not related to Spanish, a Romance language that comes from Indo-European (Aissen et al., 2017). Thus, the phonetics and phonology of Mam are very different from that of Spanish. While Spanish and Mam share 14 consonants (i.e., /p, b, m, w, t, d, n, ɾ, s, l, t͡ʃ, j, k, g/), Mam includes 19 consonants that are not part of the Spanish inventory (e.g., uvular and glottal phonemes /q, q’, χ, ʔ/, retroflex consonants / t͡s, t͡s’, s/, apico-post-alveolar consonants / t͡ʃ, t͡ʃ’, ʃ /; England, 2011; Goldstein, 2000). Spanish includes four consonants that are not part of the Mam inventory (i.e., /f, r, ɲ, x/; Goldstein, 2000). This results in a very different accent from Spanish-influenced English. Because of the higher prevalence of Spanish speakers in the US (13% of the population speaks Spanish at home; Thompson, 2021), the familiarity of Mam-accented English is very low relative to Spanish-accented English.

To summarize, the above evidence indicates that the listener is influenced by the speaker’s accent, their familiarity with an accent, and biasing effects of speaker characteristics (e.g., knowledge of client’s history, physical appearance). These factors, along with a lack of resources (i.e., phonemic inventories of lesser-known languages), may be contributing to the misdiagnoses seen with bilingual children. This leads to the current study question that asks about the impact of the speaker’s linguistic background (i.e., familiar, and unfamiliar accent) on global intelligibility assessments. We expect an impact of the visual images of the Spanish-English bilingual and Mam-English bilingual children, who appear similar but sound different from each other, such that Mam-English biligual children will be judged more harshly on the speech intelligibility judgments than Spanish-English bilingual children.

2.4.4. Training Can Overcome These Problems

The previous sections demonstrate that there are many influences on the listener based on their background, experience, exposure, and implicit biases regarding children of different ages, language abilities and linguistic backgrounds. Additionally, listeners are biased towards what they expect to hear (e.g., Kleinschmidt & Jaeger, 2015). Training and awareness may be a way to overcome the limitations of these influences. SLPs are highly trained listeners and the academic and clinical training they receive has an impact on their skills. This subsection describes the impact of training with inexperienced listeners as well as SLPs.

Experimental training has been shown to improve inexperienced listener's skills. Ellis and Beltyukova (2008) found that when lay listeners were given feedback training, they improved their ability to perform word identification tasks while listening to disordered speech. Eakins (1969) also showed the value of experimental training. They trained naïve listeners ($N = 15$) by having them listen to audio recordings of slow-played monosyllabic words produced by a man, a woman, and a child, while reading a script of words. The trained group performed significantly better than the untrained control group ($N = 20$) on a final test of orthographic transcription of distorted monosyllabic words. In a different study, explicit training was found to have great effects compared to passive training (Borrie et al., 2012). Borrie and colleagues (2012) found that when listeners were familiarized with disordered speech by coupling it with written feedback of intended targets (i.e., explicit training), there were greater subsequent gains in intelligibility scores of dysarthric speech compared to listeners who were simply familiarized with dysarthric speech (i.e., passive training). Thus, the particular type of training provided to inexperienced listeners has an influence on the gains seen.

With respect to SLP training, education in perceptual skills begins early due to the importance of these skills for clinical purposes in both assessment (i.e., detecting sound errors) and treatment (i.e., sound elicitation techniques). Students studying communication sciences and disorders take courses, such as clinical phonetics, which aim to improve auditory-perceptual abilities for detailed phonetic transcriptions and ratings (Howard & Heselwood, 2002; Lohmander et al, 2021). SLPs continue to sharpen their ability to hear and recognize phonetic detail with experience and exposure (Munson et al., 2012). However, some sound errors are still more difficult to hear than others. According to a hierarchy of difficulty in perceiving sound errors, the identification of vowel errors is more difficult than consonant errors (Allison et al., 2021; Howard & Hesselwood, 2002; Jing & Grigos, 2022). This is particularly relevant in multilingual children, compared to monolinguals, who may produce vowel differences in their native language influenced English (e.g., Ball, 2012; Peppé, 2012).

Given the advanced training SLPs receive to make high-level perceptual judgments, the current practice of collecting a speech sample and providing an intelligibility judgment provides important information about the child's speech. In other words, the current practice is valuable in addition to being practical. However, it is not known what experienced listeners are really paying attention to when they make perceptual judgments, and even trained listeners are influenced by biasing effects of speaker characteristics (Kent, 1996). So, a listener's training and knowledge about speech development and their familiarity with particular accents will influence their expectation-driven speech perception of intelligibility.

2.5. Current Dissertation

The current research investigates the extent to which impressionistic judgments of speech intelligibility are influenced by nonspeech factors, including a child's age and language abilities, as well as the listener's familiarity with an accent. This study will address the question of how we might further improve practitioner training by identifying the sources of variability in perceptual speech judgments. In turn, results will inform policy around best practices for assessment and broaden the capacity of SLPs to provide adequate services to diverse language communities.

To address the study's aim of characterizing the underlying perception factors that influence SLP's subjective intelligibility judgements, baseline information is established for speech intelligibility with a lab-based intelligibility score (LIS), and comprehensibility with a comprehensibility rating measure (CR) in the child speaker groups (i.e., younger and older school-age children in: typically developing monolingual, children with speech sound disorder, typically developing Spanish-English bilingual, typically developing Mam-English bilingual). These two measures provide a point of comparison with our more subjective measure of interest (i.e., GIA). With that information in hand (see Chapter 4), we ask four research questions.

First, do expert (i.e., SLP) listener's impressionistic intelligibility judgments (i.e., Global Intelligibility Assessments [GIA]) of children's speech differ from those of lay listeners across age and speaker groups (see Chapter 5)? We expect that SLPs are better than lay-listeners at identifying errors and hearing subtle changes in children's speech at different ages and linguistic backgrounds because they are already practiced (e.g., Monsen, 1983). We predict expert listeners will differentiate between the child speaker groups to a greater extent than lay listeners in the following ways: younger children's speech will be judged as less intelligible than older children's speech (Flipsen, 2006; Munson et al., 2010) and speech sound disordered and accented speech will be judged as less intelligible than speech produced by monolingual English children without SSD (e.g., Hustad, 2012; Lousada et al., 2014; McLeod et al., 2013).

Second, does accent familiarity impact expert and lay listeners' GIAs (Chapter 5)? An objective of this study was to test the working hypothesis that an unfamiliar accent (e.g., Mam-influenced speech) will be judged as less intelligible than a familiar accent (i.e., Spanish-influenced speech) (e.g., Derwing & Munro, 1997). This question was motivated by the fact that English learners are more likely than native speakers to be misidentified as having speech and language impairments (Artiles et al., 2002; Counts, et al., 2018; Sullivan, 2011). Under-served children who speak lesser-known languages (e.g., the Mayan language, Mam, targeted in the present study) are the most vulnerable to misidentification in part since intelligibility ratings are influenced by listener familiarity with the speaker's accent (Park, 2020).

Third, to what extent does the speech of the children influence expert and lay listeners' intelligibility judgments (see Chapter 5)? To answer this research question, we examine the relationship between global intelligibility assessments (GIA) and an experimentally controlled, laboratory-based measure of intelligibility (LIS) with the prediction that these measures will be correlated (Ishikawa et al., 2021; Stipancic et al., 2016). Additionally, we examine the relationship between global intelligibility assessments (GIA) and a rating of comprehensibility (CR) to determine whether GIA was more aligned with intelligibility (LIS) or comprehensibility (CR). Further, due to their training, expert listener's judgments are expected to be more reliable than lay listener's judgments (i.e., correlate more highly with the more objective lab-based intelligibility score [LIS] derived from an orthographic transcription task) (Brunnegård et al., 2009; Munson et al., 2012).

Fourth, does a child's language ability influence expert listeners' intelligibility judgments (see Chapter 6)? Here we operationalize language ability as the linguistic complexity demonstrated in two areas that are developing in school-age children's language: 1) utterance length, as measured by mean length of utterance in morphemes (MLUm) and 2) the diversity of vocabulary used, as measured by the number of different words (NDW). The objective was to test the working hypothesis that as children's linguistic complexity increases (i.e., greater MLUm and NDW), ratings of intelligibility also increase since children with more robust language skills will have more robust phonological systems (e.g., Edwards et al, 2004; Gathercole et al., 1999).

CHAPTER III

METHODS

To answer the research questions, speech stimuli were gathered from child speakers in four speaker groups and perceiver data were collected from two listener groups. This chapter describes the participants, the procedures for eliciting speech stimuli, how the audiovisual files were prepared, and the perceiver tasks for the experiments discussed in Chapters 4-6.

3.1. Participants

3.1.1. Child Speakers

Thirty school-age children (18 female) participated as speakers in the study. There were four speaker groups: 1) monolingual English-speaking, typically developing (*n* = 8, 4 in each age group), and 2) monolingual English-speaking, with speech sound disorder (*n* = 6[1], 2 in the younger group and 4 in the older group; below average score on the articulation subtest of the Diagnostic Evaluation of Articulation and Phonology), 3) bilingual, Spanish-English typically developing (*n* = 8, 4 in each age group), and 4) bilingual, Mam-English typically developing (*n* = 8, 4 in each age group). Each of the four speaker groups were divided into two age groups. The average age for all children in the younger group was 6 years, 1 month (range = 61 to 70 months); it was 7 years, 8 months in the older group (range = 93 to 105 months). There was slight variation in mean

[1] We aimed to recruit eight speakers in each of the groups but had difficulty finding speakers for the speech sound disorder group, perhaps due to concerns related to the COVID-19 pandemic.

ages in each of the speaker groups (younger group: monolingual typically developing *M* = 71.5 months, speech sound disorder xi = 73.5 months, Spanish-English bilingual *M* = 72.25 months, Mam-English bilingual *M* = 75.5 months; older group: monolingual typically developing *M* = 98.5 months, speech sound disorder *M* = 95.25 months, Spanish-English bilingual *M* = 98.5 months, Mam-English bilingual *M* = 102.25 months). For the monolingual groups, parents reported English as the children's only and first/native language. The English dialect was Standard American inflected by the back-vowel fronting typical of the West Coast. The Spanish-English bilingual children were all sequential learners of English except one simultaneous learner. The Mam-English bilingual children, who spoke the Todos Santos Cuchumatán dialect of Mam, were all sequential learners of English and had exposure to Spanish, but none of them were proficient speakers of Spanish.

Children were recruited from the Eugene, Oregon area in one of four ways: (a) word-of-mouth through a network of contacts and fliers distributed at summer camps; (b) University of Oregon's (UO) developmental science database, Team Duckling; (c) community liaisons to the Mam speaking community; or (d) UO's Early Dual Language Development Lab's database of Spanish-speaking families. An established community advisory board (i.e., interpreter, English Language Development teacher, family liaison) assisted with the recruitment of Mam-speaking families. A bilingual Spanish-English research assistant helped with recruitment of Spanish-speaking families.

As per parent report, the racial distribution in both the monolingual typically developing, and speech sound disordered groups was 100% White. In both bilingual groups the distribution was 100% Hispanic, although this is a problematic term for the

Mam population since they are an indigenous group to Guatemala and southern Mexico (Marcus, 2020). Since socioeconomic status is positively correlated with caregiver education (Davis-Kean, 2005), this information was collected. In the typically developing monolingual group 50% of caregivers had advanced degrees (i.e., masters, PhD, or MD) and the other 50% had a college degree (i.e., bachelors or associates). In the speech sound disordered monolingual group 50% of caregivers had advanced degrees (i.e., masters, PhD, or MD), 17% had a college degree, 17% had some college, and 17% had a high-school degree. In the Spanish-English bilingual group, 13% of caregivers had advanced degrees (i.e., masters), 13% had a college degree, 50% had not finished high school, and 26% did not report their level of education. Parents of the Mam-English bilingual children were originally from Guatemala where access to schooling was limited. Most of these parents had either completed some primary school or had no formal education, although the father of one child completed 8th grade in Guatemala.

The children's hearing was screened at 1000 Hz, 2000 Hz and 4000 Hz at a threshold of 25 dB SPL. Two participants did not pass the hearing screening (i.e., one frequency missed in one ear), but because their task performance was consistent with other participants, their data were included in the study. Speech and language development was determined in the monolingual English groups with the tools summarized in Table 2. Standardized assessment tools included the articulation subtest of the Diagnostic Evaluation of Articulation and Phonology (DEAP, Dodd et al., 2002) and the Core Language Score on the Clinical Evaluation of Language Fundamentals (CELF-5; Wiig et al., 2013). Children in the typically developing group fell in the average range on the DEAP and CELF-5. Inclusion in the speech sound disordered group was determined by below average scores on the DEAP (i.e., indicating delayed articulation) as well as average scores on the CELF-5 (i.e., indicating average expressive and receptive language skills). Of the six children in the SSD group, one child came into the lab with a confirmed SSD, two had previously been in speech therapy for an articulation delay, and three had no previous diagnosis or speech therapy. Articulation errors produced by children in the speech sound disorder group included substitutions and distortions of /ɹ/, /ð/, and /θ/ as well as one child with a frontal lisp.

All parent/guardians completed the Intelligibility in Context Scale ([ICS], McLeod et al., 2012) to help determine speech development. The ICS is a 7-item questionnaire that asks parents to rate the degree to which their child's speech is understood by different communication partners on a 5-point scale. Parents of the monolingual English children completed a written form and parents of the bilingual children were verbally asked the questions by an assistant in their native language. Normative values for English speaking children specify an average ICS score of at least 4.3 by the age of 5 years, 5 months (McLeod et al., 2015).

Table 2. Monolingual group's speech and language means and standard deviations.

Measure	Typically Developing		Speech Sound Disorder	
	M	*SD*	*M*	*SD*
DEAP	11.13	1.36	2.5	2.07
CELF-5 CLS	114.38	13.05	116.83	16.02
ICS	4.43	.49	4.38	.70

Note. DEAP = Diagnostic Evaluation of Articulation and Phonology, scaled score average range is 8-12. CELF-5 = Clinical Evaluation of Language Fundamentals, standard score average range is 85-115. ICS = Intelligibility in Context Scale, 1-5 range.

Parents of the bilingual children were interviewed in their native language using a custom bilingualism questionnaire (Appendix A). They were asked about their child's languages of exposure, onset of exposure, amount of exposure and use, and proficiency (Byers-Heinlein et al., 2019; Fabiano-Smith & Barlow, 2010; Hammer et al., 2012). Parents were asked to describe their child's schedule on a typical day and weekend days to account for all linguistic interactions (i.e., activity, individuals involved, language typically used during that activity). The number of hours the child was exposed to daily in each language was determined as "input." The same method was used for "output." Overall percentages were calculated for input and output. Table 3 presents a summary of this information and ICS means and standard deviations.

Table 3. Typically developing bilingual group's speech and language means and standard deviations.

Measure	Spanish-English		Mam-English	
	M	*SD*	*M*	*SD*
% English Input/Output	57.94	17.54	53	5.88
% Native Language Input/Output	42.06	17.54	47	5.33
Years of Exposure to English	4	1.93	2.63	1.06
ICS	4.68	.46	4.98	.05

Note. ICS = Intelligibility in Context Scale, 1-5 range.

Typical speech and language development was determined in the bilingual groups by parent report and an informal evaluation in both of their languages. English was evaluated with nonstandard administration of the DEAP and the CELF-5; subtests were administered as with the monolingual English-speaking children; however, standardized scores were not calculated. Articulation probes in Mexican Spanish (Bilingual Speechie, n.d.) were elicited with picture prompts and analyzed by referencing the Spanish Phonemic Inventory and facts on Spanish Phonology (ASHA, n.d.). Typical articulation development in the Todos Santos Cuchumatán dialect of Mam was determined with a custom screening tool (Appendix B) that was developed based on Mam consonant and vowel inventories

(England, 2017; U. Canger, personal communication, November 3, 2021) and an analysis of Mayan child phonetic inventories (Pye et al., 2017).

The children's typical language development in Spanish or Mam was determined based on 1) parent report, 2) parent rating of language proficiency, 3) referral of Mam-speaking children by their English Language Development teacher, and 4) discussion with the interpreters regarding two language samples. Parents rated their child's language proficiency for both English and their native language as 3 or 4 on a 0-4 scale where 0 = child cannot speak the indicated language at all and 4 = child has native-like proficiency in the language. The language samples were elicited with *The Ball Mystery* and *Baseball Troubles* from the School-age Language Assessment Measures (SLAM, Crowley & Baigorri, 2015). Prior to telling the narrative, the children were asked standard questions about the story pictures. Spanish versions of the questions were accessed from the SLAM website and a Mam interpreter translated the questions for administration to the Mam-English bilingual children.

3.1.2. Expert Listeners (Speech-Language Pathologists)

Forty-two[2] licensed SLPs (5 male, 1 nonbinary; age 26-70 years; Median age: 35.5 years) with a broad range of experience working with children (2-40 years; Median:

---

[2] Prior to recruitment, it was determined that forty listeners in each perceiver group (i.e., expert and lay) would provide at least 92% power to detect a large effect size with a correlation of r^2=0.5 (Cohen, 1988) and a 5% significance level. Data from 43 participants were collected, however one participant did not complete all the tasks correctly, so this data was not included in the analyses.

9 years) were recruited from across the US as expert listeners. Two SLPs were bilingual, non-native speakers of English (i.e., native German and Thai speakers). Expert listeners were asked to rate their exposure to Spanish and to Mam on a 7-point Likert scale where: 1 = “never heard the language” and 7 = “hear it spoken daily.” Their exposure to the Spanish language was reported as moderate (*M* = 4.31; *SD* = 1.65) on a 1-7 scale. Their reported exposure to the Mam language on the same scale was reported as minimal (*M* = 1.24; *SD* = .53). This confirmed that Mam was an unfamiliar language to the SLPs relative to Spanish.

An IRB-approved recruitment script was emailed to the primary investigators network of SLP contacts and posted on the Oregon Speech and Hearing Association’s (OSHA) website, newsletter, and social media. Additionally, the recruitment script was posted on the primary investigator’s social media accounts and on three relevant listservs: two special interest groups of the American Speech and Hearing Association (SIG 19, Speech Science and SIG 14 Cultural and Linguistic Diversity) and the Cognitive Development Society. SLPs were compensated for their participation with $20 gift cards and a 10% discount on the virtual OSHA annual conference which provided Continuing Education Units.

3.1.3. Lay Listeners

Forty-one[3] native speakers of English, aged age 18;9 to 23;5 years, including 22 females (54%) and 19 males (46%) with no self-reported speech, language, or hearing impairments, were recruited as lay listeners from the UO's Psychology and Linguistics Human Subjects Pool. All lay listeners received course credit for their participation. These participants completed a background survey that asked about their level of exposure to children (i.e., low, medium, high). Twelve reported a high exposure to children in roles such as camp counselor, childcare provider, or having younger siblings or nieces/nephews. Fourteen reported medium exposure to children and 15 reported low exposure in similar roles. Lay listener's exposure to the Spanish language was reported as moderate ($M = 3.74$; $SD = 1.32$) on a 1-7 Likert scale (1 = "never heard the language" and 7 = "hear it spoken daily"). Their reported exposure to the Mam language on the same scale was reported as minimal ($M = 1.04$; $SD = .21$). These means and standard deviations are slightly lower than those reported by the expert listeners but continue to provide evidence that the perceivers in both listener groups had some familiarity with Spanish and next to no familiarity with Mam.

3.2. Materials and Procedures

3.2.1. Speech and Language Elicitation

Note: Speech data were collected from children during the COVID-19 pandemic while a mask mandate was in place. Following UO policy, children wore masks except during

---

[3] Data from 42 participants were collected, however one participant did not complete the task correctly, so this data was not included in the analyses.

short, videotaped speech tasks. Experimenters wore masks throughout the entire protocol and added a face shield during the times children had their masks removed.

All child speakers completed three English-speaking tasks (see Table 4) in random order: 1) conversational speech, 2) structured spontaneous language samples, and 3) a sentence repetition task of 40 semantically anomalous sentences. Conversational speech and language samples were chosen since these are commonly used by SLPs to complete global intelligibility assessments (GIA) like the one in the current study. Semantically anomalous sentences (e.g., "Wide pens swim fast," Appendix E) were chosen as stimuli to reduce any contextual effects on perceivers in the experimentally controlled lab-based intelligibility scoring (LIS) and comprehensibility rating (CR) tasks (see section 3.2.3). These three speaking tasks, along with speech and language assessments, were blocked by language for the bilingual child speakers and the order of the languages was counterbalanced.

Table 4. Child speaking tasks.

Task	Elicitation materials	Recording method	Purpose
Speech sample: Conversation	Open-ended questions	Video	Global intelligibility assessments (GIA)
Language sample: Narratives	2 SLAM stories	Audio	Measures of linguistic complexity (MLUm/NDW)
Sentence repetition	40 low-predictability sentences	Audio	Lab-based intelligibility score (LIS) and comprehensibility ratings (CR)

Note. SLAM = School-Age Language Assessment Measures, MLUm = mean length of utterance in morphemes, NDW = number of different words.

Speech tasks were audio-visually recorded in a sound-attenuated (clinical-type) observation room at the University of Oregon Speech and Language Lab for all children except the Mam-English bilinguals. To eliminate the transportation burden to the lab, the Mam-English bilingual children were seen in their school district. They were recorded in a quiet office environment at the South Lane School District's Family Resource Center in Cottage Grove, Oregon.

The children's conversational speech was audio-video recorded and elicited with open-ended questions (e.g., *Tell me about your favorite video game*) (Abbeduto, 2021; Sturm & Seery, 2007). A timer was set for three minutes, and the children were prompted with questions based on the experimenter's observations of the child's interest. For example, if a child reported that they recently had a birthday party, the experimenter asked about the details of that celebration.

Two structured spontaneous language samples were audio-video recorded for each child speaker. Samples were elicited using the same materials for each child, which included six or seven picture prompts for two stories, *Dog Comes Home* and *Bunny Goes to School,* which were counterbalanced across participants. These stories were taken from the School-age Language Assessment Measures (SLAM, Crowley & Baigorri, 2015), a tool designed to elicit language samples in preschool- and school-aged children. For each story, the children were first asked standard questions provided with the SLAM materials (see Appendix C) while looking at picture cards, which had been placed in front of them by the experimenter. The questions acted to prime the child's language about the story and give them time to familiarize themselves with the storyline (Potratz et al., 2022). The questions required the participant to deduce (e.g., *Why is she in the bathtub with a white* *dog now?*), infer (e.g., *Why did she come to school?*), problem solve (e.g., *What would you do if a bunny started hopping around your school?*), predict (e.g., *What do you think will happen when the boy goes home?*), and use theory of mind (e.g., *What is the girl thinking here?*). Then, with the pictures still in front of them, the children were asked to tell the story shown in the pictures (see Appendix D for sample narratives). During language sample elicitation, the experimenters provided natural prompts, such as encouragement to begin or continue the story narrative.

Two language sample measures, mean length of utterance in morphemes (MLUm) and number of different words (NDW), were used to index each child's level of linguistic complexity for one of the experimental variables. MLUm and NDW were chosen as representative measures since they are commonly used and their values discriminate the language skills of children in different age groups (i.e., 5- and 8-year-old children; Potratz et al., 2022). To calculate these measures, audio-video files of the children's language samples were transcribed using Systematic Analysis of Language Transcripts (SALT; Miller & Iglesias, 2019) conventions to identify utterances, words, morphemes, unintelligible segments, and mazes (i.e., filled pauses, repetitions, revisions, and abandoned utterances) (Miller et al., 2019). Specifically, the samples were first segmented into C-units (i.e., one main clause and any modifiers or subordinate clauses; Loban, 1976) using C-unit segmentation rules (Miller et al., 2019). Utterances that were less than a C-unit were included (e.g., *sorry; bye*) if they were not maze behavior (i.e., false starts, repetitions, reformulations, and filled pauses, e.g., *um; uh; well*). Stereotypic closing (e.g., *the end*) and side comments were placed on special lines and were not included as part of the transcription. MLUm and NDW were automatically calculated using SALT. Averages of the two English language narratives (e.g., *Dog Comes Home* and *Bunny Goes to School*) provided by each child speaker were used for the analysis.

To assure transcription, C-unit segmentation, and coding reliability, we used a consensus procedure (as in Guo & Eisenberg, 2015; Shriberg et al., 1984). Each sample was initially transcribed and segmented by the primary investigator. Then, a trained research assistant reviewed the data while listening to the recorded language samples and reading the initial segmented transcriptions. Transcription, segmentation, or coding disagreements were identified, and then reviewed and discussed until agreement was obtained for all transcripts (e.g., Frizelle, et al., 2018).

The sentence repetition task included 40 semantically anomalous sentences that were elicited from the children for use in two perceiver tasks, the lab-based intelligibility scoring (LIS) and comprehensibility ratings (CR). Semantically anomalous, or low-predictability, sentences were chosen to decrease the listener's ability to use semantic context; they reduce the use of top-down processing, preventing the listener from inferring misperceived words based on the context (e.g., McLaughlin et al., 2018). The sentences, adapted from Stelmachowicz et al. (2000), consisted of four content words each (e.g., "Wide pens swim fast"; See Appendix E) that were selected to be within the vocabulary of children as young as 4-years-old. A female, native standard American English speaker was recorded reading the 40 sentences in two blocks of 20 in a quiet environment. The two blocks of 20 were counterbalanced when played back to the children for repetition. The children repeated each sentence following the pre-recorded adult model and were allowed to hear the sentence multiple times if they did not remember it on the first hearing.

3.2.2. Preparing Audiovisual Files for Perceiver Tasks

Video files of the conversational speech task were edited into two-minute clips using iMovie software for the global intelligibility assessments (GIA). Audio files of the sentence repetition task were cut into individual sentence files using Audacity software (Audacity Team, 2012) and were amplitude normalized to 70 dB in Praat (Boersma & Weenink, 2022). These stimuli were embedded with multi-talker babble using a custom Praat script with a signal-to-noise ratio of +5 dB. Multi-talker babble was chosen as a masker to avoid a ceiling effect on our measures (i.e., lab-based intelligibility score and comprehensibility rating) since it has properties like the speech signal and is used extensively in speech perception research. Stimuli were preceded by 400 ms of babble noise and followed with 50 ms of babble noise. Audio files were converted from wav to mp3 format for Testable (see section 3.2.3) capability using the Apple Music application.

3.2.3. Perceiver Tasks

An online platform, Testable (www.testable.org), was used for the experiments with lay and expert listeners. This platform allowed for survey questions and the presentation of video and audio stimuli with customization for participant answer selections and capabilities for randomization and blocking. After the participants agreed to the consent form and completed a background survey, each of the three perceiver tasks (see Table 5) were presented in a counterbalanced order: 1) global intelligibility assessments (GIA), 2) lab-based intelligibility scoring (LIS), and 3) comprehensibility ratings (CR). Audio and video files were randomized into four batches that each included eight video files (one from each speaker/age group) and 64 audio files for LIS and 64 different audio files for CRs. SLPs completed a post-experiment survey asking about their language background, how they define and measure intelligibility in practice, the training they have received with regards to measuring speech intelligibility, and what speech characteristics they were paying attention to when listening to the study's speech samples. All perceivers were debriefed after the experiment.

Table 5. Perceiver task instructions including speakers, and stimuli (LIS and CR included the same speakers but different sentences).

Task	Instructions	Speakers	Stimuli
Global intelligibility assessment (GIA)	Please wear earphones and participate in a quiet environment. After watching a 2-minute video of a child speaking please type the percentage (0-100) of their speech you understood and press Enter/Return.	Random subset of 8 speakers, 1 from each speaker/age group.	8 video files with 2 minutes of conversational speech.
Lab-based intelligibility score (LIS)	After listening to each 4-word sentence that is embedded in noise, please type what you hear. Type x for any word you don't understand.	8 speakers, 1 from each speaker/age group (different from speakers in the video files).	8 noise-masked, semantically anomalous sentences per speaker (n=8) for a total of 64 sentences.
Comprehensibility rating (CR)	After listening to each 4-word sentence that is embedded in noise, please rate how much you understood.	8 speakers, 1 from each speaker/age group (different from speakers in the video files).	8 noise-masked, semantically anomalous sentences per speaker (n=8) for a total of 64 sentences.

For GIA, all perceivers viewed eight two-minute videos of conversational speech and then indicated the percentage of speech they understood. The SLPs were additionally asked about the speaker's language abilities after viewing each video. For LIS, all perceivers listened to and orthographically transcribed 64 sentences that were blocked by speaker group, to ensure each perceiver heard speakers from each group, and masked with multi-talker babble to prevent a ceiling effect. The order of the sentence presentation was randomized to create eight unique lists of sentences, each consisting of eight sentences each. After hearing each utterance, the perceivers were asked to type what they heard and if they did not understand something they were asked to type an “X” in place of the word or words. The LIS was then manually calculated as the number of words correctly identified by the listeners in each utterance (0 = none through 4 = all). Omissions or additions of final consonants denoting plurality or tense were ignored, as were incorrect word order and the inclusion of additional words. Misspellings and homonyms were accepted as correct. A mean LIS for each child speaker was calculated from 80 sentences that were transcribed by individuals from both listener groups.

Interscorer reliability of LIS was completed on 13% of the data. All the transcribed data were initially scored for the LIS by the primary investigator. The transcribed data of four randomly selected speakers, one from each speaker group, were then independently rescored by a trained research assistant. The original LIS results for the same speakers were then correlated with the rescored transcription results, yielding a Pearson product-moment correlation coefficient of .99 across all speakers. This indicated a very high level of reliability for scoring accuracy (r (430) = .99, $p < .001$).

CRs were gathered from the perceivers on 64 of the noise-masked semantically anomalous sentences that were blocked by speaker. These sentences were from the same speakers but different utterances from the LIS task. After hearing each sentence, the perceivers were presented with a slider that they could move to indicate their rating (i.e., a 7-point Likert scale where: 1 = “easy to understand” and 7 = “difficult to understand”).

A mean CR for each child speaker was calculated from 80 sentences rated by individuals from both listener groups.

CHAPTER IV

BASELINE MEASURES: LAB-BASED INTELLIGIBILITY SCORES (LIS) AND COMPREHENSIBILITY RATINGS (CR)

4.1. Introduction

This chapter establishes experimentally controlled, baseline measures of intelligibility and comprehensibility of the child participants in each age and speaker group. These more objective measures will serve as reference points of comparison with the impressionistic, global intelligibility assessments (GIA) presented in Chapter 5.

Both expert and lay listener groups listened to and orthographically transcribed controlled, semantically anomalous sentences, which provided lab-based intelligibility scores (LIS) (i.e., number of content words correct). This measure was chosen since orthographic transcription measures are the gold-standard method for more objectively measuring intelligibility in various populations, such as individuals with dysphonia, children with SSD, and adults with dysarthria (Hustad, 2006; Ishikawa et al., 2021; Lousada et al, 2014; Stipancic et al., 2016). Additionally, this tool provided a quantifiable measure of the perceiver's understanding of the speech signal, allowing us to capture this construct.

We tested the performance of LIS as a measure of intelligibility to distinguish between the two age groups and between the four speaker groups. The stimuli used in this task, semantically anomalous sentences embedded in noise, were chosen intentionally to lessen the influence of language ability and the confound of context. It was hypothesized that LIS would delineate age groups such that the older group would have scores that were clearly higher than the younger group on this task (see Hustad et al, 2021) due to the continued development of the phonetic system during the early school-age years (e.g., Smit et al., 1990). Likewise, we anticipated differences between speaker groups such that monolingual English typically developing children would have scores that were higher than the other three speaker groups due to influences of speech errors in the SSD group and accent in the two bilingual groups. Specifically, a lack of familiarity with nonnative English in the perceivers could influence LIS of the bilingual speaker groups (Derwing & Munro, 1997; Munro & Derwing, 1995). Finally, it was predicted there would be no differences in LIS between the expert and lay listener groups, since this type of task is less subjective and hence would not be greatly influenced by the experience of the expert listeners.

Comprehensibility ratings were gathered from the perceivers on these same sentences for comparison (i.e., a 7-point Likert scale where: 1 = "easy to understand" and 7 = "difficult to understand"). As with the LIS measure, we predicted differences between age groups and speaker groups, with children in the older group (mean age 7;8) receiving higher comprehensibility ratings compared to children in the younger group (mean age of 6;1) and children in the monolingual English typically developing group receiving higher comprehensibility ratings compared to children in the monolingual English SSD and bilingual groups. As with LIS, a lack of familiarity with nonnative English in the perceivers could influence the CR of the bilingual speakers.

Comprehensibility rating scales, like the one we used, are an established way of examining listener's perceptions of nonnative speech (e.g., Derwing & Munro, 2009; Munro & Derwing, 1995; Munro & Derwing, 2001; Trofimovich & Isaacs, 2012). Additionally, studies that have examined speech impaired populations have shown rating scales to be effective and reliable. For example, Ishikawa and colleagues (2021), in tasks completed by inexperienced listeners, found a strong positive correlation between transcription-based and ratings-based intelligibility measures with adult and child participants with dysphonia. Stipancic and colleagues (2016), who also used inexperienced listeners, found a positive correlation between orthographic transcription of sentences and ratings on a visual analog scale using speech from adults with mild dysarthria (i.e., due to Parkinson's disease and multiple sclerosis) and healthy controls. These studies demonstrate the reliability of rating scales and positive outcomes even with untrained listeners, both of which give confidence in rating scales as a good measure to compare to our impressionistic assessments (GIA).

To reiterate, the goal of this chapter was to establish baseline measures with intelligibility scores and comprehensibility ratings of the child speakers. Using these measures, we had a means for comparison when evaluating the GIA in the subsequent chapter.

4.2. Analyses and Results

Analyses were completed using R Studio (Rstudio Team, 2020), a companion program to R (R Core Team, 2020). R packages were used for data management (tidyr; Wickham & Girlich, 2022), analysis (lme4; Bates et al., 2015, sjstats; Lüdecke, 2021), and visualization (ggplot2; Wickham, 2016, smplot; Seung, 2022). First, we analyzed effects on LIS (i.e., LIS was the dependent variable), with a linear mixed-effects model that included fixed effects of the listener group, speaker group, and speaker age group, with individual speakers, individual listeners, and data batch number (i.e., 1-4, as described in the Methods chapter) as random effects. To test for a predicted interaction between age group on LIS, the full model was compared to a reduced model that removed age group. A similar model comparison was completed for speaker group. Second, the effects on CR (i.e., CR was the dependent variable) were analyzed with a linear mixed-effects model that had listener group, speaker group, and speaker age group as fixed effects and individual speakers, individual listeners, and data batch number as random effects. Reduced models removed age group and speaker group for comparison to the full model. In all cases, interaction terms were removed from the models when not significant. Partial eta-squared was calculated to determine effect sizes, where .01 = a small effect, .06 = a medium effect, and .14 = a large effect (Field, 2013).

4.2.1. Lab-Based Intelligibility Scores (LIS)

As shown in Table 6, the linear mixed-effects model, where LIS was the dependent variable, indicated no main effect of listener group on LIS. This shows that both expert and lay listener groups performed the LIS task similarly. There was a significant main effect of age group on LIS, showing that the older group was more intelligible than the younger group. There was also a significant main effect of speaker group for both the monolingual English SSD and Mam-English bilingual groups. So, speakers in the Mam-English bilingual and monolingual English SSD groups were harder to understand relative to the monolingual English typically developing group.

Table 6. Linear mixed-effects model results with LIS as the dependent variable.

	Model summary			
	β	SE	t	p
Fixed effects				
Intercept	2.06	.22	9.37	<.001***
Listener Group	-.03	.07	-.39	.697
Age Group	.63	.19	3.37	.002**
Speaker Group Mam	-1.02	.25	-4.04	<.001***
Speaker Group Spanish	-.40	.25	-1.58	.126
Speaker Group SSD	-.77	.27	-2.83	.009**
	s^2			
Random effects				
Listener	.09			
Speaker	.25			
Batch	.06			

Note. Reference for Listener Group is expert listeners, for Speaker Group is monolingual English typically developing, for Age Group is younger group, ***$p < .001$, **$p < .01$, *$p < .05$.

The overall mean LIS by child speaker ranged from .58 to 3.13 out of a possible 4.0. The speech of the children in the younger group (mean age of 6;1) was scored lower ($M = 1.53$, $SD = 1.23$) than the speech of the children in the older group (mean age of 7;8, $M = 2.13$, $SD = 1.33$). This is illustrated in the box plot in Figure 2. Model comparisons between the full model and a reduced model with age group removed, indicated a significant relationship between age group and LIS with a small effect size ($\chi^2 = 9.52$, $p = .002$, $\eta p^2 = .002$), confirming the full model result.

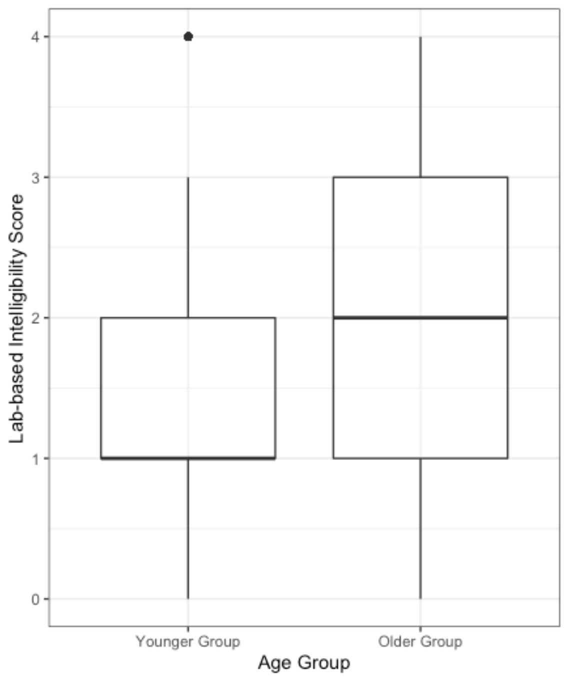


Figure 2. Lab-based intelligibility scores (LIS) by age group.

Mean lab-based intelligibility scores (LIS) along with standard deviations are shown in Table 7 by speaker group, where speakers in the monolingual English typically developing group received the highest LIS, followed by the typically developing Spanish-English bilingual group, the monolingual English SSD group, and the lowest scoring were speakers in the typically developing Mam-English bilingual group. Model comparisons between the full model and a reduced model with speaker group removed, indicated a significant relationship between speaker group and LIS with a small effect size (χ^2 = 14.05, p = .003, ηp^2 = .003), again confirming the full model result. The visual presentation of the mean LIS in the four speaker groups is shown in Figure 3.

Table 7. Mean lab-based intelligibility scores (LIS) for each speaker group.

Speaker Group	*M*	*SD*
TD Monolingual	2.36	1.29
SSD Monolingual	1.64	1.25
TD Spanish-English Bilingual	1.97	1.30
TD Mam-English Bilingual	1.36	1.20

Note. TD = typically developing, SSD = speech sound disorder.

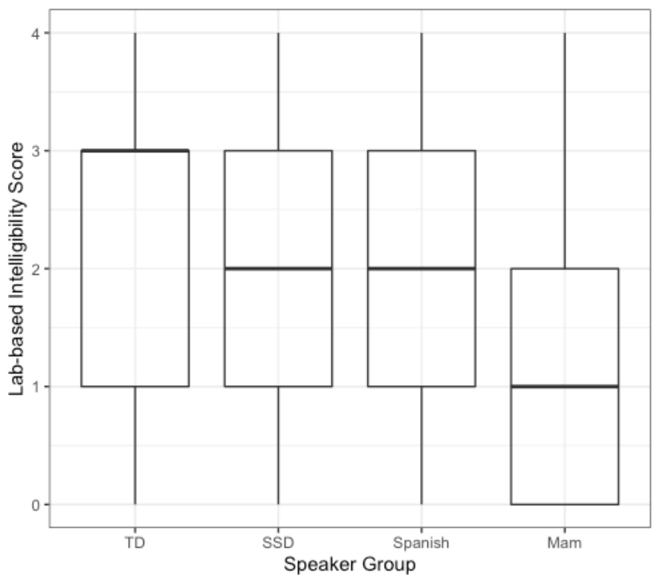


Figure 3. Lab-based intelligibility score for each speaker group.

Note. TD = monolingual English typically developing, SSD = monolingual English speech sound disorder, Spanish = typically developing Spanish-English bilingual, Mam = typically developing Mam-English bilingual.

4.2.2 Comprehensibility Ratings (CR)

Complete full model results for CR as the dependent variable are presented in Table 8. As with the LIS analysis above, there were significant main effects of age group and speaker groups for the SSD and Mam-English bilingual groups, but no effect of listener group on CR. Both expert and lay listener groups rated the children's speech similarly, with the younger speaker group being less comprehensible than the older speaker group. Children in the Mam-English bilingual and SSD groups were harder to understand relative to children in the typically developing group.

Table 8. Linear mixed-effects model results with comprehensibility rating (CR) as the dependent variable.

	Model summary			
	β	SE	t	p
Fixed effects				
Intercept	4.17	.31	13.43	<.001***
Listener Group	-.23	.17	-1.38	.172
Age Group	.83	.26	3.23	.003**
Speaker Group Mam	-1.31	.35	-3.74	<.001***
Speaker Group Spanish	-.41	.35	-1.18	.250
Speaker Group SSD	-1.26	.38	-3.32	.003**
	s^2			
Random effects				
Listener	.53			
Speaker	.47			
Batch	.02			

Note. Reference for Listener Group is expert listeners, for Speaker Group is monolingual English typically developing, for Age Group is younger group, ***$p < .001$, **$p < .01$, *$p < .05$.

The mean CR by child speaker ranged from 1.76 to 6.34 out of a possible 7.0. The speech of the children in the younger group (mean age of 6;1) was rated lower (*M* = 3.34, *SD* = 1.91) than the speech of the children in the older group (mean age of 7;8, *M* = 4.16, *SD* = 1.98), as illustrated in the box plot in Figure 4. Model comparisons between the full model and a reduced model with age group removed, indicated a significant relationship between age group and CR with a small effect size ($\chi^2 = 8.84$, $p = .003$, $\eta p^2 = .002$), confirming the full model result.

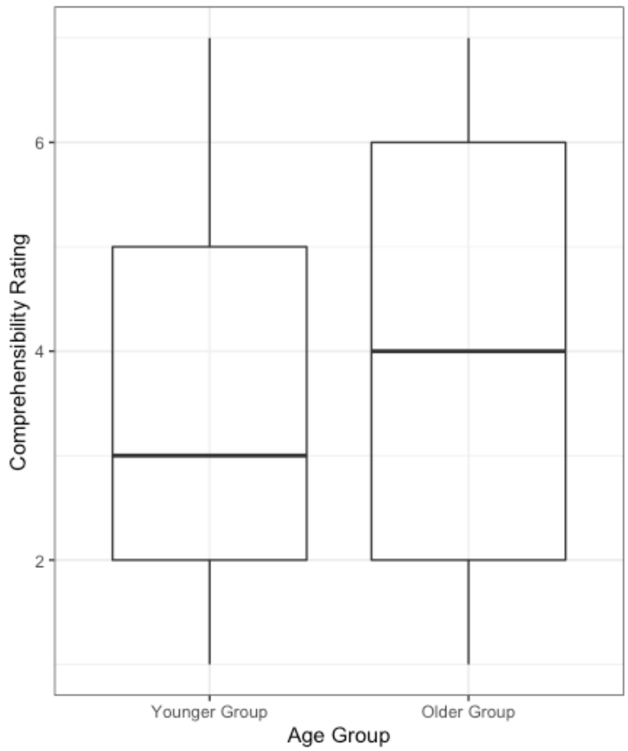


Figure 4. Comprehensibility ratings (1-7) by age group.

The means and standard deviations of CR for each speaker group are shown in Table 9. These values show the same pattern seen above for LIS: children in the typically developing group were rated as having the highest comprehensibility, followed by children in the Spanish-English bilingual group, then children in the SSD group, and the lowest rated were children in the Mam-English bilingual group. Model comparisons between the full model and a reduced model with speaker group removed, indicated a significant relationship between speaker group and CR with a small effect size (χ^2 = 14.62, p = .002, ηp^2 = .004), again confirming the full model result. Figure 5 shows the mean CRs for each speaker group in a box plot.

Table 9. Overall means and standard deviations for comprehensibility ratings (CR) from 1-7 of each speaker group.

Speaker Group	*M*	*SD*
TD Monolingual	4.48	1.96
SSD Monolingual	3.25	1.79
TD Spanish-English Bilingual	4.06	2.00
TD Mam-English Bilingual	3.20	1.90

Note. TD = typically developing, SSD = speech sound disorder.

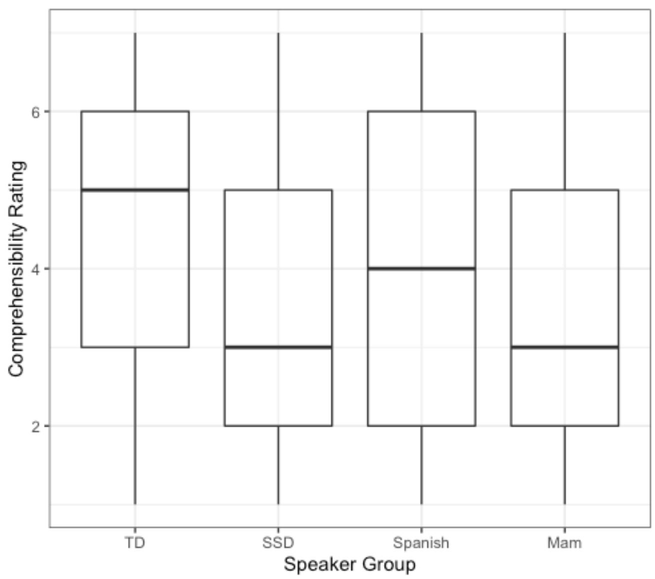


Figure 5. Comprehensibility ratings (1-7) by speaker group.

Note. TD = monolingual English typically developing, SSD = monolingual English speech sound disorder, Spanish = typically developing Spanish-English bilingual, Mam = typically developing Mam-English bilingual.

4.3. General Discussion

The analyses in this chapter, which examined a lab-based intelligibility score (LIS) and a comprehensibility rating (CR), generated similar results for both expert and lay listener groups; no main effect or interaction was found for listener group. Since the experience and prior academic and clinical training of the expert listeners did not have an impact on their scores and ratings, we can assume a level of objectivity and reliability of these measures. In this case, we can also conclude that inexperienced listeners are able to rate the comprehensibility of child speakers as capably as expert listeners (e.g., Hashemi Hosseniabad et al., 2021). Establishing LIS and CR as reliable measures is important since the findings from these measures are used as a point of comparison to the impressionistic, global intelligibility assessment (GIA) measure in Chapter 5.

Statistical tests showed differences between age groups in their intelligibility (LIS) and comprehensibility (CR), as was expected. Older children, due to the development of speech sounds, are more easily understood. Likewise, LIS and CR both showed significant differences between the disordered speech of the children in the monolingual English SSD group and the speech of children in the monolingual English typically developing group even though their errors were relatively mild. LIS and CR both showed significant differences between the speech of the children in the Mam-English bilingual group and the speech of children in the monolingual English typically developing group. The speech of the Spanish-English bilingual children was essentially scored as intelligible and rated as comprehensible as the speech of the children in the English monolingual typically developing group, so there was no influence of a Spanish accent on these two measures.

What does it mean that LIS and CR had similar results? This could indicate that the two measures assess the same construct or are "tapping into the same perceptual phenomenon" (Stipancic et al., 2016). As such, the measures could potentially be used interchangeably, as has been suggested by several researchers (e.g., Hustad, 2006; Ishikawa et al., 2021; Stipancic et al., 2016). Ishikawa and colleagues (2021), in their study of patients with dysphonic speech, concluded that "the rating-based measurement has a potential substitution for the transcription-based analysis." Stipancic and colleagues (2016), in their study of participants with mild dysarthria, made a similar conclusion, highlighting the fact that rating scales are less time consuming. While this appears to be a free pass for busy clinicians to use rating scales, the type of speech impairment and the severity of the speech issue need to be considered. More research is needed in this area, along with more nuanced suggestions that incorporate the type of disorder and severity.

It should be noted that several of the bilingual child speakers had significant difficulty repeating the semantically anomalous sentences that were the stimuli in the LIS and CR tasks. While the children were allowed multiple attempts and the prompt was repeated for them to improve their repetition, there were still many errors (e.g., mispronunciation of parts of the target sentences and word substitutions). This was particularly the case for the children in the Mam-English bilingual group. Recall that the Mam-English bilingual group had less exposure to English (years of English exposure: $M = 2.63$, $SD = 1.06$) compared to the Spanish-English bilingual group (years of English exposure: $M = 4$, $SD = 1.93$). So, it is possible that their limited English exposure impacted their ability to perform the sentence repetition task (e.g., Meir et al., 2016). As such, there was likely an impact of the materials themselves on the speaker group outcomes due to the impact of the errors on the range of mean LIS and mean CR (see Hashemi Hosseinabad et al., 2021).

In the next chapter, the effect of listener experience on impressionistic, global intelligibility assessments (GIA), frequently used by practicing SLPs, is examined. The lab-based intelligibility score (LIS) and comprehensibility rating (CR) measures investigated in the current chapter are used as a basis of comparison to GIA.

CHAPTER V

EFFECT OF LISTENER EXPERIENCE ON GLOBAL INTELLIGIBILITY ASSESSMENTS (GIA)

5.1. Introduction

With the experimentally controlled, baseline measures of lab-based intelligibility scores (LIS) and comprehensibility ratings (CR) in hand, we now turn to the heart of the dissertation which asks about the integrity of the impressionistic judgments that SLPs make when evaluating children for speech sound disorder (SSD). Here we refer to those judgments as global intelligibility assessments (GIA). The purpose of this chapter is multifold. First, we examine the GIAs of our two listener groups to determine whether listener expertise makes a difference in these judgments. Second, we examine GIA across the two age groups to see if GIA is significantly different between age groups like it was for LIS and CR. Third, we examine GIA across the four speaker groups to see if listener's GIAs were sensitive to differences between those groups. Finally, we examine the relationships between GIA and the baseline measures from Chapter 4, LIS and CR, with correlational analyses. The correlational analyses were repeated with data split by listener group to examine differences between the reliability of expert versus lay listener GIAs.

We begin by addressing the first research question which asked whether there was an effect of listener group (i.e., expert vs lay) on impressionistic, global intelligibility assessments (GIA) of children's speech. Here we defined expert listeners as licensed SLPs who, as part of their clinical training, have received instruction in listening to children's speech and making intelligibility judgments. Lay listeners were untrained, and in this case, we had university undergraduate students fulfill that role. The objective was to test the working hypothesis that expert listeners, due to their training and experience, would differentiate between the child speaker groups (i.e., older and younger school-age children in monolingual English typically developing, monolingual English speech sound disorder, typically developing Spanish-English bilingual, and typically developing Mam-English bilingual groups) to a greater extent than lay listeners. We predicted that expert listeners would rate older children's speech as more intelligible than younger children's speech (Flipsen, 2006; Munson et al., 2010) to a greater extent compared to lay listeners due to their training and exposure to developmental differences. Further, regarding the children's linguistic background (i.e., monolingual / bilingual, disordered), we expected expert listeners to rate speech sound disordered and accented speech as less intelligible than speech produced by their typically developing peers (e.g., Hustad, 2012; Lousada et al., 2014; McLeod et al., 2013), again to a greater extent than lay listeners.

Differences have been found between expert and lay listener's ability to make impressionistic intelligibility judgments. SLP's academic and clinical training and experience appear to have an impact: in most cases SLPs are superior judges and more reliable judges of children's articulatory skills compared to lay listeners. SLP's superiority as listeners differs however depending on the listening task and the speaker population. That is, greater differences are seen between listener experience groups, contingent upon the linguistic complexity of the speech samples and the capabilities of the speakers. For example, greater listener experience effects occur when identifying gradations of correctness (Klein et al., 2012) and with highly unintelligible speakers producing linguistically complex sentences (Monsen, 1983). Differences between inexperienced listeners and listeners who have experience with a particular speaker group have also been documented (e.g., hearing impaired: Monsen, 1983; child client with misarticulated /r/: Wolfe et al., 2003). Given this evidence, we expected expert listeners, due to their training and experience, to better able to distinguish between age and speaker groups with GIAs.

The second research question asked about speaker group differences and whether there was an influence of accent familiarity on expert and lay listeners' GIAs. As discussed in Chapter 2, Section 2.4.3., the objective was to test the working hypothesis that children with an unfamiliar accent (i.e., Mam-influenced English) would be judged as less intelligible than those with a familiar accent (i.e., Spanish-influenced English). In other words, we predicted that child speakers with an unfamiliar accent would be rated as being both less intelligible and comprehensible than those with a familiar accent (e.g., Derwing & Munro, 1997). We made this prediction since an unfamiliar, complex mix of non-standard features, such as altered syllable stress patterns or non-standard segments, have been shown to reduce intelligibility in speakers of English as a second language (Bent, Bradlow, & Smith, 2007; Zielinski, 2006). As such, we expected differences between the bilingual groups for two reasons: 1) the familiarity of a Spanish-influenced English gave those children a "pass" and 2) the visual/auditory mismatch related to the Mam-English speaking children who appeared like the Spanish-English bilingual children in the videos but sounded different.

The third research question asked about the extent to which the speech signal influenced listeners' intelligibility judgments. To answer this question, we analyzed the relationships between global intelligibility assessments (GIA) and an experimentally controlled, laboratory-based measure of intelligibility (i.e., LIS), as well as between GIA and comprehensibility ratings (CR). We hypothesized that if GIA and LIS were correlated, then GIA would be aligned with the speech signal since LIS provides a more objective value without the confounds of language and contextual cues. Additionally, if GIA and CR were correlated, then GIA would be aligned with listener perceptions of how much they understood the speakers (Ishikawa et al., 2021; Munro & Derwing, 2001; Stipancic et al., 2016).

Further, we tested the working hypothesis that expert listener's GIAs would be more reliable than lay listener's GIAs, and thus correlate more highly with the lab-based intelligibility score (LIS) (Brunnegård et al., 2009; Munson et al., 2012). Experienced listeners are known to make judgments that are more reliable (i.e., more consistent), more valid (i.e., more closely related to acoustic characteristics of sounds), and more precise (e.g., identify gradations of correctness; greater sensitivity and specificity) than inexperienced listeners (Brunnegård et al., 2009; Hashemi Hosseinabad et al., 2021; Klein et al, 2012; McFarlane et al., 1991; Munson et al., 2012). Thus, we predicted that we would see differences between the listener groups in terms of the level of correlation between GIA and LIS.

To summarize, this chapter examines the effect of listener expertise on GIA. We test the hypothesis that, due to their training and experience, expert's ratings would differ from lay listener's ratings for both the speaker's age and linguistic background (i.e., monolingual English typically developing, monolingual English SSD, or typically developing bilingual). We predicted that listener experience would be an influential variable in these perceptual judgments tasks such that expert listeners would be better able to distinguish between age and speaker groups compared to untrained, lay listeners.

Finally, based on the SLP survey (Table 1) we anticipate that GIAs are more related to listener understanding, or comprehensibility (via CR) than to LIS, which is more reflective of the speech signal. The results were expected to further our understanding of the differences that occur with training and how we can further enhance that training.

5.2. Analyses and Results

Analyses were completed using R Studio (Rstudio Team, 2020), a companion program to R (R Core Team, 2020). R packages were used for data management (tidyr; Wickham & Girlich, 2022), analysis (lme4; Bates et al., 2015, sjstats; Lüdecke, 2021), and visualization (ggplot2; Wickham, 2016, smplot; Seung, 2022). We analyzed the effects on GIA with a linear mixed-effects model that included fixed effects of the listener group, speaker group, and speaker age group, with individual speakers, individual listeners, and data batch number (i.e., 1-4, as described in the Methods chapter) as random effects. To test for a predicted interaction between listener status (i.e., expert versus lay) on GIA, the full model was compared to a reduced model that removed listener group. Similar model comparisons were completed for age group and speaker group. The same models were run split by listener group to determine whether there were differences between expert and lay GIAs. Interaction terms were removed from the models when not significant. Partial eta-squared was calculated to determine effect sizes, where .01 = a small effect, .06 = a medium effect, and .14 = a large effect (Field, 2013).

To examine GIA further, we analyzed it in relation to LIS and CR. To test the strength of the overall relationship between GIA and LIS, Pearson product-moment correlation coefficient analyses were completed. Then for comparison, the data were split by listener group and correlational analyses were again completed. Similar analyses tested the strength of the relationship between GIAs and CRs, overall and with data split by listener group.

5.2.1. Effect of Listener Group

Table 10 presents the full model results of the linear mixed-effects model that examined the effects of listener group, age group, and speaker group on GIA. Interactions were removed since they were not significant. There was a significant main effect of listener group on GIA such that expert listeners rated children's speech as being more intelligible (*M* = 90.95, *SD*=12.06) compared to lay listeners (*M* = 85.22, *SD* = 16.61), as predicted. There was also a significant main effect of speaker group on GIA, but only for the Mam-English bilingual speaker group whose children received much lower values on the GIA compared to the children in the monolingual English typically developing group.

Table 10. Linear mixed-effects model results with GIA as the dependent variable.

	Model summary			
	β	SE	*t*	*p*
Fixed effects				
Intercept	85.18	2.46	34.70	<.001***
Listener Group	-5.78	2.00	-2.89	.005**
Age Group	1.93	1.75	1.11	.278
Speaker Group Mam	-7.49	2.41	-3.11	.004**
Speaker Group Spanish	.53	2.41	.22	.828
Speaker Group SSD	-2.88	2.58	-1.12	.275
	s^2			
Random effects				
Listener	69.12			
Speaker	17.91			
Batch	2.29			

Note. Reference for Listener Group is expert listeners, for Speaker Group is monolingual English typically developing, for Age Group is younger group, ***$p < .001$, **$p < .01$, *$p < .05$.

Model comparisons between the linear mixed-effects full model and a reduced model with listener group (i.e., expert and lay) removed, indicated a significant relationship between listener group and GIA with a small effect size ($\chi^2 = 7.94$, $p = .005$, $\eta p^2 = .014$), confirming the full model result. Figure 6 illustrates the differences between expert and lay listener GIAs. A visual comparison of the expert and lay listener interquartile ranges in the box plot shows a smaller range for the expert listeners. This indicates less variability in SLP's GIAs, as they may be using common criteria to make their assessments.

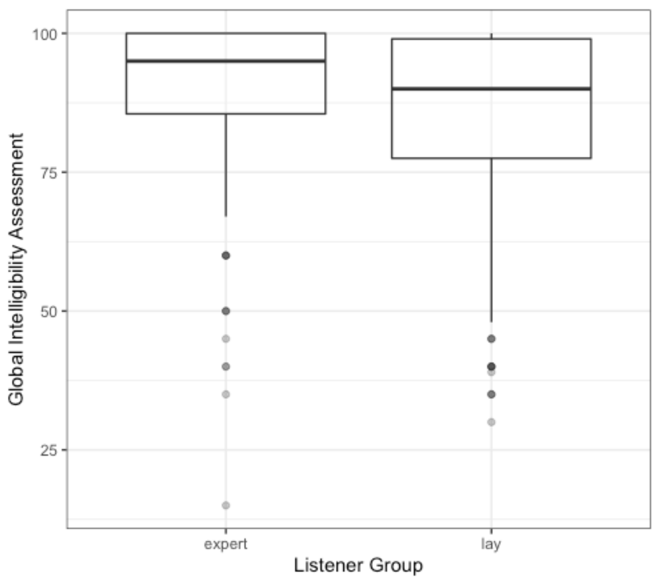


Figure 6. Box plot of expert and lay listener's Global Intelligibility Assessments (GIA).

5.2.2. Effect of Age Group

Overall, descriptive statistics indicate that the children in the younger group (mean age of 6;1) were rated on the GIA as being less intelligible ($M = 86.87$, $SD = 15.35$) than the children in the older group (mean age of 7;8, $M = 89.35$, $SD = 14.07$), as would be expected given developmental speech changes. However, model comparisons between the linear mixed-effects full model and a reduced model with age group removed, was not significant ($\chi^2 = 1.17$, $p = .279$, $\eta p^2 = .002$), confirming the full model result (see Table 10). So, surprisingly, the GIAs did not distinguish between the two age groups in the study.

5.2.3. Effect of Speaker Group

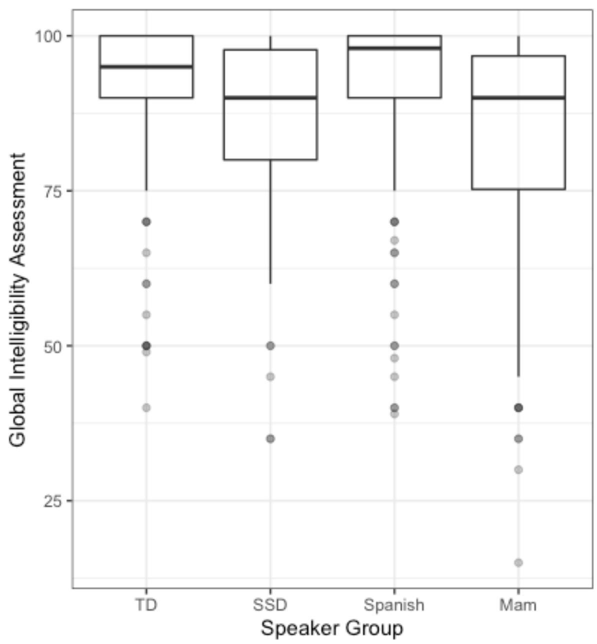
Note that the Mam-English bilingual children were audio and video recorded in a different environment from the other three groups. While unlikely, this may have caused some differences in the global intelligibility assessments (GIA) that were obtained using video samples.

Descriptive statistics, presented in Table 11, show the overall means and standard deviations of the global intelligibility assessments (GIA) for each of the four speaker groups. Children in the typically developing group received the highest GIA, followed by children in the Spanish-English bilingual group, then the speech sound disordered group, and the children in the Mam-English bilingual group received the lowest GIA. The box plots in Figure 7 illustrate the differences in GIA across the four speaker groups. Model comparisons between the linear mixed-effects full model and a reduced model with speaker group removed, was significant with a small effect size ($\chi^2 = 11.29$, $p = .010$, $\eta p^2 = .024$), indicating that the linguistic background of the child speakers did have an influence on GIA.

Table 11. Means and standard deviations of global intelligibility assessments (GIA) for each speaker group.

Speaker Group	*M*	*SD*
TD Monolingual	90.72	12.50
SSD Monolingual	87.18	13.59
TD Spanish-English Bilingual	91.35	13.71
TD Mam-English Bilingual	83.22	17.45

Note. TD = typically developing, SSD = speech sound disorder.



Note. TD = monolingual English typically developing, SSD = monolingual English speech sound disorder, Spanish = typically developing Spanish-English bilingual, Mam = typically developing Mam-English bilingual.

Figure 7. Global intelligibility assessments by speaker group.

5.2.4. Data Split by Listener Group

Despite the lack of interactions in the above models, we ran the same models split by listener group to determine whether there were differences between expert and lay GIAs. This was completed due to our fundamental interest in the effect of experience and since there was a main effect of the listener group variable in the full model. The means, standard deviations, and ranges of the global intelligibility assessments (GIA) are shown in Table 12 for each of the four speaker groups, this time by listener group. For all four speaker groups, the expert listener's mean GIA was higher than lay listener's mean GIA. The lower standard deviations for the expert listener GIA indicate less variation in ratings compared to lay listeners.

Table 12. Means, standard deviations, and ranges of global intelligibility assessments (GIA) for each speaker group by listener group.

Speaker Group	Lay Listener GIA ($N = 41$)		Expert Listener GIA ($N = 42$)	
	M (SD)	Range	*M (SD)*	Range
TD Monolingual	88.68 (14.53)	49-100	92.73 (9.77)	40-100
SSD Monolingual	85.00 (14.67)	35-100	89.31 (12.15)	35-100
TD Spanish-English Bilingual	88.44 (16.62)	30-100	94.22 (9.29)	40-100
TD Mam-English Bilingual	78.76 (18.62)	35-100	87.58 (15.09)	15-100

Note. TD = typically developing; SSD = speech sound disorder

Table 13 presents the full model results of the linear mixed-effects analysis that examined the effects of age and speaker group and their interactions on expert listener's GIA. Table 14 presents the results of a similar model but with lay listener's data. There were several things to note when analyzing these results. First, the variance (s^2) in all three of the random effects (i.e., individual speaker, individual listener, and batch) is much greater in the lay listener results compared to the expert listeners as noted above. Second, regarding the speaker groups, it was the Mam-English bilingual and the monolingual SSD groups that had significant main effects in both expert and lay listener groups. Third, there were significant interactions between age group and the SSD group for both expert and lay listener groups. Fourth, the interaction between age group and the Mam-English bilingual speaker group was only significant for the expert listener group, not for the lay listener group. A visual presentation of these data showing GIA by speaker group and age group is shown in box plots in Figure 8, with expert listener data on the left and lay listener data on the right.

Table 13. Linear mixed-effects model results for expert listeners only with GIA as the dependent variable.

	Model summary			
	β	SE	t	p
Fixed effects				
Intercept	94.10	2.17	43.41	<.001***
Age Group	-2.67	2.75	-.97	.338
Speaker Group Mam	-10.16	2.74	-3.71	<.001***
Speaker Group Spanish	-1.93	2.74	-.71	.486
Speaker Group SSD	-8.19	3.04	-2.69	.014*
AgexSpeaker Group Mam	10.07	3.88	2.58	.014*
AgexSpeaker Group SP	6.50	3.89	1.67	.104
AgexSpeaker Group SSD	9.49	4.10	2.32	.029*
	s^2			
Random effects				
Listener	4.00			
Speaker	7.07			
Batch	2.53			

Note. Reference for Speaker Group is typically developing monolingual English and for Age Group is younger group, ***p < .001, **p < .01, *p < .05

Table 14. Linear mixed-effects model results for lay listeners only with GIA as the dependent variable.

	Model summary			
	β	SE	t	p
Fixed effects				
Intercept	91.91	3.88	23.67	<.001***
Age Group	-6.19	3.84	-1.61	.117
Speaker Group Mam	-12.60	3.94	-3.45	.002**
Speaker Group Spanish	-4.82	3.94	-1.22	.232
Speaker Group SSD	-10.26	4.49	-2.29	.033*
AgexSpeaker Group Mam	6.99	5.51	1.27	.215
AgexSpeaker Group SP	8.80	5.51	1.60	.121
AgexSpeaker Group SSD	12.95	5.91	2.19	.039*
	s^2			
Random effects				
Listener	84.76			
Speaker	18.38			
Batch	20.98			

Note. Reference for Speaker Group is typically developing monolingual English and for Age Group is younger group, ***p < .001, **p < .01, *p < .05

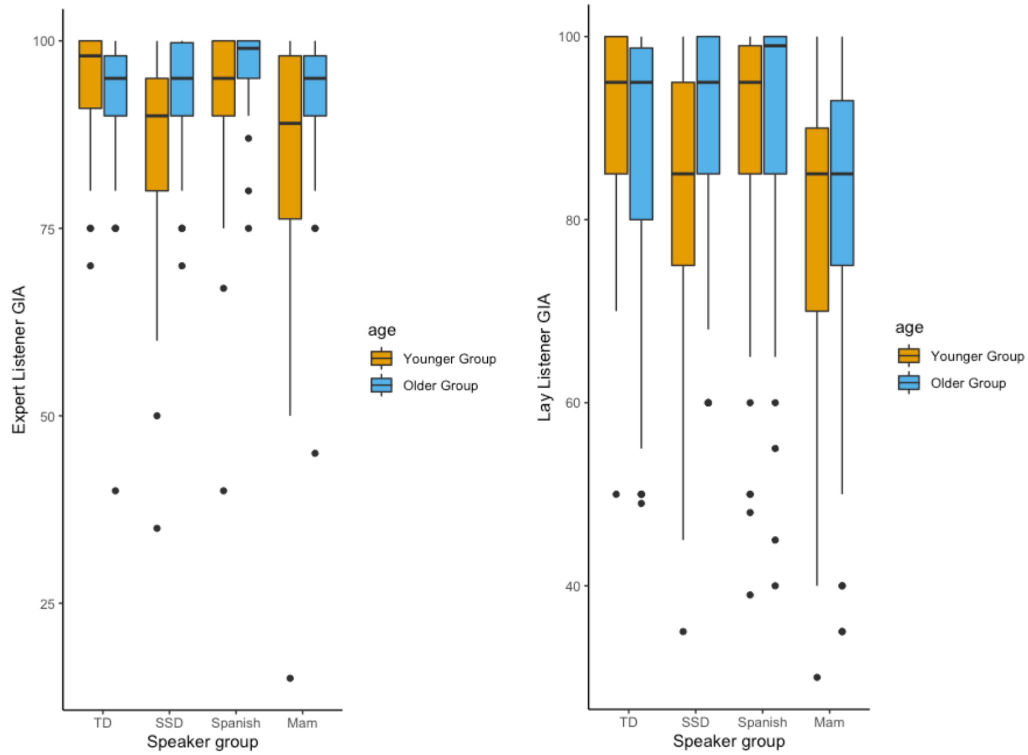


Figure 8. Box plots of GIA by speaker group and age for expert listeners (left) and lay listeners (right).

Note. TD = monolingual English typically developing, SSD = monolingual English speech sound disorder, Spanish = typically developing Spanish-English bilingual, Mam = typically developing Mam-English bilingual.

5.2.5. Correlation between GIA and LIS

To test the relationship between GIA and the more objective intelligibility measure, LIS, a Pearson product-moment correlation coefficient analysis was performed. This showed an overall statistically significant positive correlation (r (28) = .53, p = .003) (see Figure 9). This correlation remained significant even when the data were divided by listener group (see Figure 10). The data from the expert listeners appears to have a slightly weaker correlation between GIA and LIS (r (28) = .41, p = .024), compared to the lay listener data (r (28) = .46, p = .010), however a test of the difference between the two correlations was not significant (z = -.271, p = .787; Soper, 2022). These positive correlation coefficients suggest that as LIS increases, so too does GIA, regardless of the listener group.

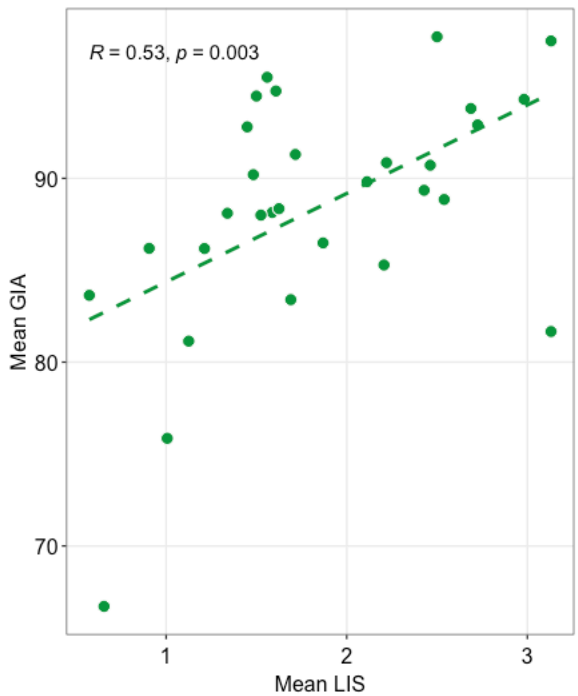


Figure 9. Correlation plot showing the overall relationship between mean global intelligibility assessments (GIA) and mean lab-based intelligibility scores (LIS).

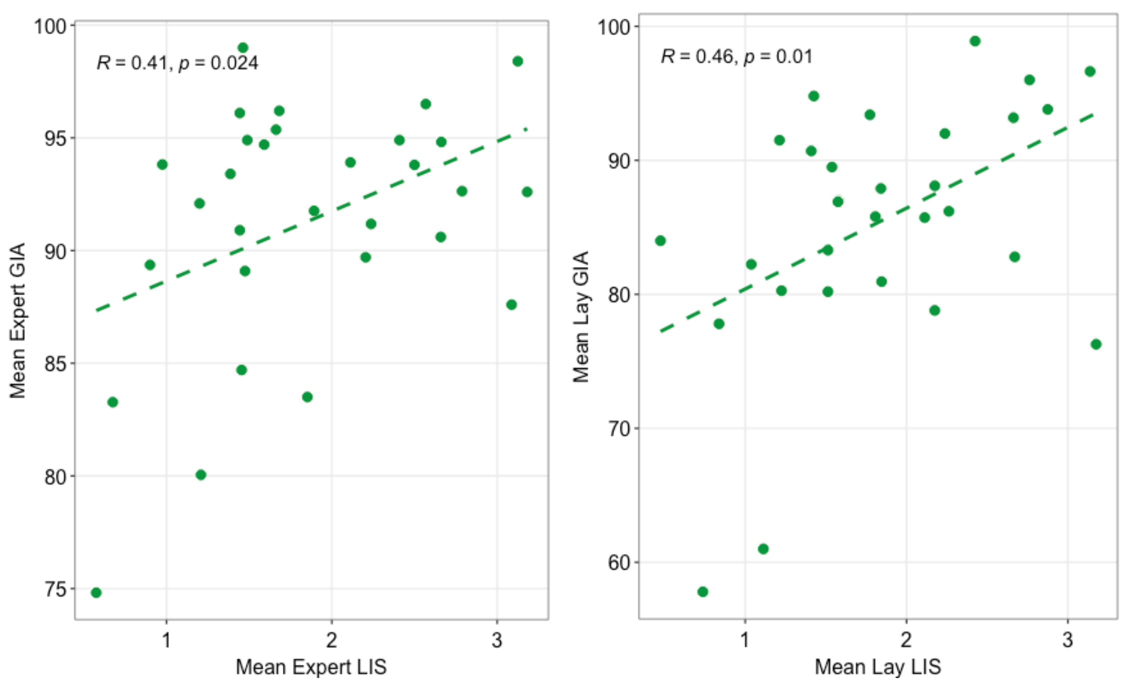


Figure 10. Correlation plots showing the relationship between mean global intelligibility assessments (GIA) and mean lab-based intelligibility scores (LIS) for expert listeners (left) and lay listeners (right).

5.2.6. Correlation between GIA and CR

A Pearson product-moment correlation coefficient analysis was performed between GIA and CR and showed an overall statistically significant positive correlation (r (28) = .62, p < .001) (Figure 11). As with the analysis of GIA and LIS above, this correlation remained significant even when the data were divided by listener group (see Figure 12). The data from the expert listeners showed comparable results (r (28) = .52, p = .003) to the data from the lay listeners (r (28) = .55, p = .002); with no significant difference between the correlations (z = -.184, p = .854; Soper, 2022). These positive correlation coefficients suggest that a greater GIA is related to a greater CR. This was again the case for both expert and lay listener group data.

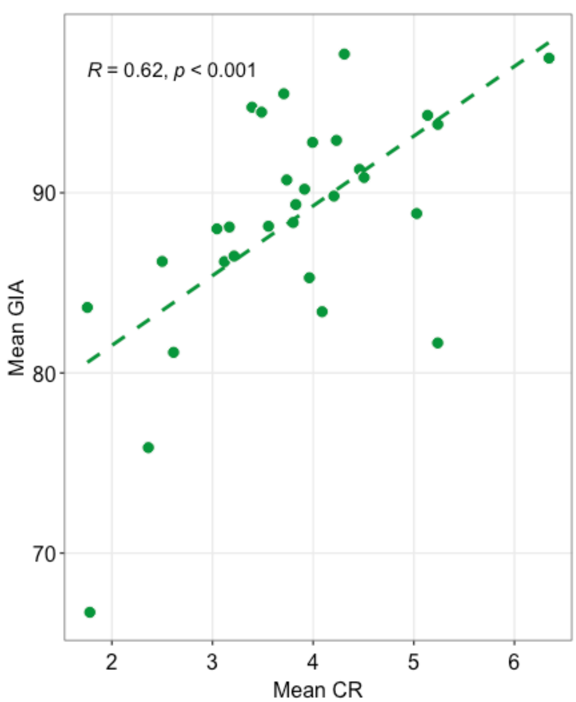


Figure 11. Correlation plot showing the overall relationship between mean global intelligibility assessments (GIA) and mean comprehensibility ratings (CR).

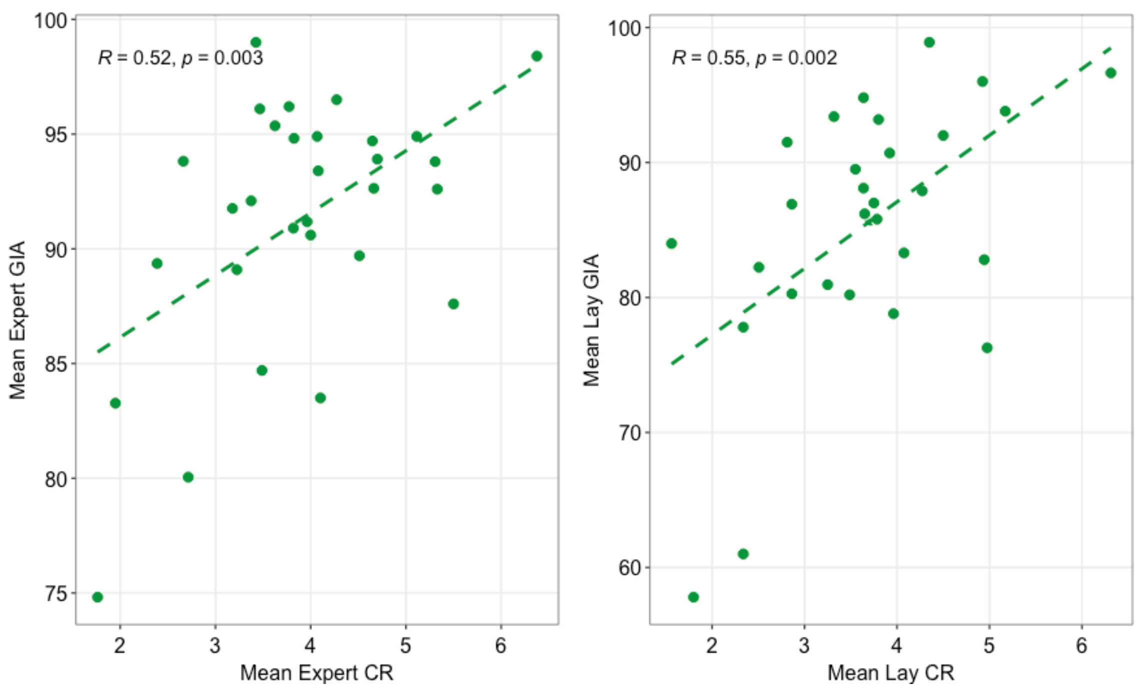


Figure 12. Correlation plots showing the relationship between mean global intelligibility assessments (GIA) and mean comprehensibility ratings (CR) for expert listeners (left) and lay listeners (right).

5.3. General Discussion

In this chapter, we investigated the impressionistic global intelligibility assessments (GIA) that were made by lay and expert listeners of younger and older school-age children in four speaker groups (i.e., typically developing monolingual English, speech sound disorder monolingual English, typically developing Spanish-English bilingual, and typically developing Mam-English bilingual) to track the impact of the expert's training and experience. Previous studies have shown differences in perceptual skills between experienced and inexperienced listeners (Hashemi Hosseinabad et al., 2021; McFarlane et al., 1991; Munson et al., 2012). Thus, the prediction here was for an interaction between listener group and the speaker variables of age and linguistic background. But that is not what was found; expert listener's GIAs were different from lay listener's GIAs and were more consistent in their rating of speakers, but they did not differentiate between speaker groups any more than lay listeners and in some cases less so. The lack of interactions underscores the subjectiveness of this measure and the impact of listener expectations on GIA.

Recall that in Chapter 4 there was not an effect of listener group on either LIS or CR; expert and lay listeners performed these two tasks in a similar fashion. This reinforced the idea that LIS and CR were more objective and reliable measures, and that there were fewer influencing factors on the perceivers. However, in the current chapter analyses, we found that expert and lay listeners provided different GIA values from each other. This suggests that the two listener groups were being influenced by something different. This could be related to the fact that GIA utilized audio-visual files of spontaneous speech samples. Experienced listeners made judgments that were more reliable (i.e., less variability in Figure 6) compared to lay listeners, perhaps because they were applying a common criterion to inform their GIA ratings (Hashemi Hosseinabad et al., 2021; McFarlane et al., 1991; Munson et al., 2012). In other words, the clinical training and experience that the SLPs have received led to a stronger agreement when judging speech intelligibility (e.g., Borrie et al., 2012; Ellis & Beltyukova, 2008) but experts were still influenced by something other than speech.

Our results also showed no significant difference in GIA of the two age groups of children in the study. This was contrary to our prediction that differences would occur between age groups due to developmental changes. Additionally, this finding was especially surprising since both LIS and CR did indeed show differences between the two age groups, which further draws into question what may be influencing GIA. One explanation for this result is that listeners were perhaps making accommodations for age differences in the GIAs of the children. Since listeners had visual cues from the videos, they may have adjusted their expectations for how understandable the children's speech should be for their approximate age. In other words, seeing a younger-appearing child producing age-expected speech errors (e.g., gliding of /ɹ/) may have elicited an equal rating to an older-appearing child who did not produce those types of errors.

The current study also revealed that of the different speaker groups, GIA only showed a significant difference for the Mam-English bilingual group, relative to the typically developing monolingual English group. So, we can speculate that the listener's familiarity with an accent had an influence on GIA, such that the speech of children in the Mam-English bilingual group was judged more harshly than that of the monolingual English group. This was not the case with the speech of children in the Spanish-English bilingual group, which was a more familiar language to the listeners. We know that the listeners had a very low level of familiarity with the Mam language, and it appears that this lack of familiarity (Adank et al., 2009; Derwing & Munro, 1997; Gass & Varonis, 1984; Huang et al., 2016) may have influenced the judgments of Mam-influenced speech. What was it about the Mam-English bilingual children's speech that produced lower GIAs? One reason may be due to speech differences secondary to cross-linguistic influences such as language interaction (Fabiano-Smith & Goldstein, 2010) resulting in speech that was less understood. Additionally, listeners may have placed a negative social evaluation on the Mam-English bilingual children resulting in lower intelligibility ratings as "minority accents are often disparaged" (Munro et al., 2006).

Were there influences on the listener based on visual information provided in the videos, or that speakers had an accent that did not match the listener's expectations?

While we do not have measures of solely visual information, we can speculate that the visual aspect of the experiment may have influenced the listeners as well. The Mam-English bilingual children had a similar physical appearance to the Spanish-English bilingual children, but their speech did not sound the same. This mismatch may have led to lower intelligibility judgments (McLaughlin & Van Engen, 2020; Van Engen & Peelle, 2014). Finally, were there influences of the language abilities of the Mam-English bilingual children? We follow up on this question in Chapter 6.

Second, why was it that GIA did not identify the SSD group from the typically developing group as was the case in both the LIS and CR analyses? This may be partially accounted for by the fact that the SSD group was only mildly impaired, producing errors that would be considered typical for younger children (e.g., substitutions and distortions of /s/, /ɹ/, /ð/, and /θ/). Perhaps GIA is not a sensitive enough measure to identify mild SSD on its own.

The results of the correlational analyses showed a strong relationship between both GIA and LIS, as well as between GIA and CR. This implies that the impressionistic GIA has elements of both intelligibility, reflected by speech signal features, and comprehensibility, reflected by listener-focused levels of understanding. The connection between GIA and comprehensibility is perhaps unsurprising since the majority of SLPs define intelligibility in terms of understanding (see Table 1 survey responses). The fact that global intelligibility assessments are aligned with comprehensibility is viewed here as a positive outcome since this reflects the functional value of this assessment.

When the data were split between listener groups and correlational analyses were run, expert listener's mean GIA correlated with their mean LIS to the same extent as lay listener's. This was also the case when comparing the correlational analyses of GIA and CR across lay and expert listener's data. These findings imply that the training and experience of the SLPs did not make a difference for GIA in this context. While it was predicted that expert GIAs would be more valid and more precise than inexperienced listeners (Brunnegård et al., 2009; Klein et al, 2012), this was not the case.

The results in this chapter begin to show the listener expectation influences on impressionistic, global intelligibility assessments (GIA) and the alignment of GIA with both comprehensibility and intelligibility. Results point to the importance of training to improve the reliability and validity of GIAs for clinicians assessing children with SSD. This is especially true for those SLPs working with children with unfamiliar accents like the Mam-English bilingual children in our study. In the next chapter, we delve deeper into what GIA is by examining how language ability may be playing a role with this measure.

CHAPTER VI

EFFECT OF LANGUAGE ABILITY ON EXPERT'S GLOBAL INTELLIGIBILITY ASSESSMENTS

6.1. Introduction

The analyses in this chapter investigated the influence of the children's language ability on expert listener's global intelligibility assessments (GIA). The goal of this chapter was to better understand what GIA measures. To achieve this goal, we examined the extent to which the expert listener's GIA were focused on language-based factors. Here we operationalized language ability as linguistic complexity measured by syntactic complexity and lexical diversity (i.e., mean length of utterance in morphemes [MLUm] and number of different words [NDW]) drawn from spontaneous speech samples (i.e., SLAM stories described in Chapter 3). These two measures were chosen since they are established measures that are commonly used in both monolingual and multilingual speakers, and they have been shown to be valid for discerning between developmental levels in school-age children. MLUm measures the length and complexity of a child's utterances (i.e., syntactic complexity) and NDW measures the variety of vocabulary used by the child (i.e., lexical diversity). It was predicted that GIA would increase with the linguistic complexity of the speech sample, demonstrated with increases in MLUm and NDW. This prediction was based on evidence that children with better language skills are also likely to have better speech skills (e.g., Edwards et al., 2004; Gathercole et al, 1999).

This final research question has become even more interesting considering the results of the previous chapter. First, consider the results in Chapter 5 indicating that both intelligibility and comprehensibility are part of the GIA. Given this, one might expect that language factors, such as syntactic complexity and lexical diversity are important, potential top-down influences on the listeners. Second, we have already seen that listeners apparently make some sort of accommodation for age differences and yet one fully expects that language will change with age. In other words, the null results of age group on GIA (Chapter 5) suggest an accommodation in the face of true speech differences (i.e., as measured by LIS) and, we are likely to find, true language differences, to be determined in the present chapter. However, we have also seen the listeners apparently do not make a similar accommodation for speakers whose first language is unfamiliar to the listener (i.e., the Mam-English bilingual participants). This begs the question, why was GIA lower only for the Mam-English bilingual group? Were listeners in fact assessing the Mam-English bilingual group on their speech signal features alone? Recall that in Chapter 4, we found that LIS (i.e., the objective measure) was already lower in Mam-English bilingual speakers than in Spanish-English bilingual speakers; Spanish-English children were not different from monolingual typically developing children, but the Mam-English bilingual children were. This suggests a true effect of Mam on the children's intelligibility. There was a similar effect of Mam on comprehensibility (i.e., as measured by CR in Chapter 4). And we saw that this effect repeated with GIA in Chapter 5.

Critically, we do not know if the impact of Mam was truly a speech effect or if it was a language effect on GIA by delving further into why there was an effect of accent familiarity on GIA. We are testing whether the effect might be based on language rather than on speech. Here, we investigate two measures of linguistic ability to determine whether the differences are upheld. In this chapter, we test for an effect of language abilities on the global intelligibility assessments to see if differences were due to language ability.

To summarize, the goal of this chapter was to examine the effects of linguistic complexity on the global intelligibility assessments made by expert listeners. We do this by first presenting the linguistic measure results by age group and speaker group and then we present the statistical analyses that test the effects of MLUm and NDW on GIA. If the influences of linguistic ability are present, training SLPs on impressionistic judgments can more explicitly specify the importance of attending to these areas.

6.2. Analyses and Results

Analyses were completed using R Studio (Rstudio Team, 2020), a companion program to R (R Core Team, 2020). R packages were used for data management (tidyr; Wickham & Girlich, 2022), analysis (lme4; Bates et al., 2015, sjstats; Lüdecke, 2021), and visualization (ggplot2; Wickham, 2016, smplot; Seung, 2022). First, we analyzed the effects of age and speaker group on MLUm and NDW separately, with linear mixed-effects models that both included fixed effects of the speaker group and age group, with individual speakers and data batch number (i.e., 1-4, as described in the Methods chapter) as random effects. Next, to test the predictive strength of these two language measures (i.e., MLUm and NDW), we analyzed for their effects on expert listener GIA (i.e., GIA was the dependent variable), with a linear mixed-effects model that included fixed effects of MLUm and NDW, with individual speakers, and data batch number (i.e., 1-4, as described in the Methods chapter) as random effects. Finally, a model was run with age group entered as a control variable to remove the variance in expert GIA due to this factor and test whether language ability, as measured by MLUm and NDW, might explain the effect of speaker group on GIA found in Chapter 5. Interactions were removed when not significant.

6.2.1. Linguistic Ability by Age Group and by Speaker Group

The mean MLUm and mean NDW were higher in the older group (mean age of 7;8, MLUm, $M = 8.37$, $SD = 1.55$; NDW, $M = 40.93$, $SD = 12.11$) compared to the younger group (mean age of 6;1) (MLUm, $M = 7.60$, $SD = 1.31$; NDW, $M = 38.07$, $SD = 12.15$) indicating stronger linguistic skills as expected due to developmental changes. Table 15 displays the means and standard deviations for the two language sample measures by age and speaker group. This is also illustrated in the box plots of MLUm (Figure 13) and NDW (Figure 14). Here we see two results in the bilingual groups which were unexpected given that these children were acquiring English as a second language. First, the Spanish-English bilingual children had the highest mean MLUm of all four speaker groups, in both younger and older age groups. Second, the highest mean NDW was in the older Mam-English bilingual group. The children in the speech sound disorder group performed similarly to the Mam-English bilingual children on MLUm and had the lowest NDW of all the speaker groups, in both age groups.

Table 15. Means, standard deviations, and ranges for each of the language sample measures by speaker group and age group (younger group: mean age of 6;1, older group: mean age of 7;8).

Age	Measure	TD Monolingual		SSD Monolingual		TD Spanish-English Bilingual		TD Mam-English Bilingual	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Younger	MLUm	7.72	.86	7.08	.47	7.98	1.32	6.92	1.47
	NDW	38.34	6.62	31.32	3.00	41.14	14.60	35.93	14.90
Older	MLUm	8.95	.71	7.37	1.94	9.38	.52	7.94	1.42
	NDW	42.42	11.65	31.64	7.85	41.00	6.20	47.29	10.84

Note. MLUm = mean length of utterance in morphemes, NDW = number of different words, SSD = speech sound disorder, TD = typically developing.

While the descriptive data appear to show differences between age groups and speaker groups, neither of the linear mixed-effects models that tested the effect of speaker group and age group on MLUm, and separately on NDW, showed significance of age (with the younger group as the reference) or speaker group (with the typically developing monolingual English speakers as the reference) (MLUm: Mam speaker group, $t = -6.66$, $p = .781$; Spanish speaker group, $t = -1.13$, $p = .872$; SSD speaker group, $t = -7.96$, $p = .648$; age group, $t = 5.50$, $p = .583$; NDW: Mam speaker group, $t = .793$, $p = .728$; Spanish speaker group, $t = .450$, $p = .770$; SSD speaker group, $t = -5.47$, $p = .664$; age group, $t = 3.04$, $p = .257$). Since there were very few speakers per comparison group (i.e., $n = 4$) in these analyses, the power was probably not adequate to show a significant result. Despite this, the data on MLUm and NDW presented in Table 15 and Figures 13 and 14 do give a sense of how the various groups (i.e., age and speaker) differed.

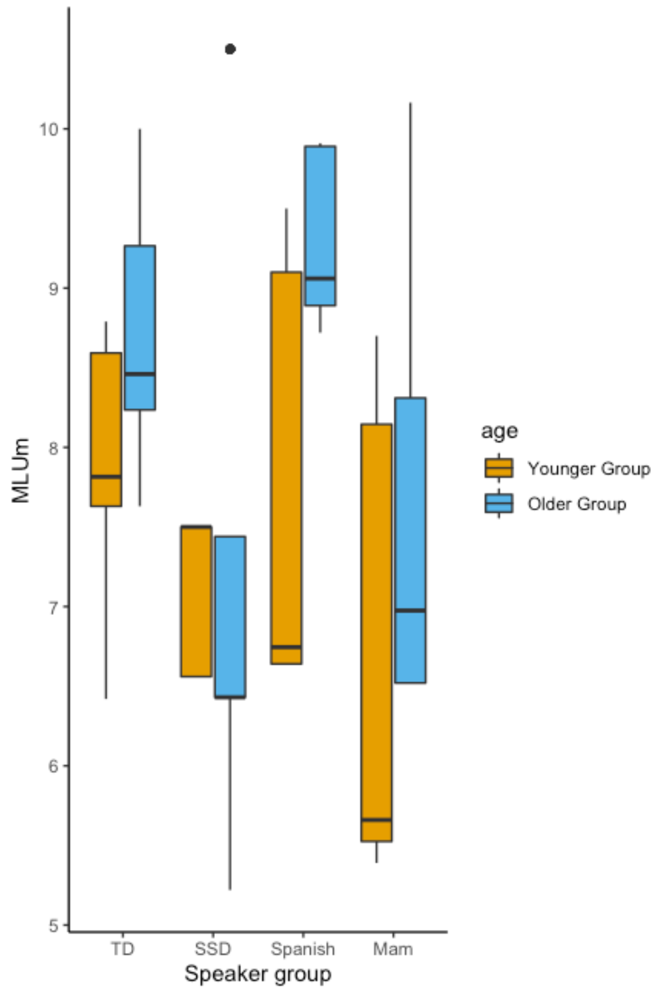


Figure 13. Mean length of utterance in morphemes (MLUm) by age and speaker group.

Note. TD = monolingual English typically developing, SSD = monolingual English speech sound disorder, Spanish = typically developing Spanish-English bilingual, Mam = typically developing Mam-English bilingual.

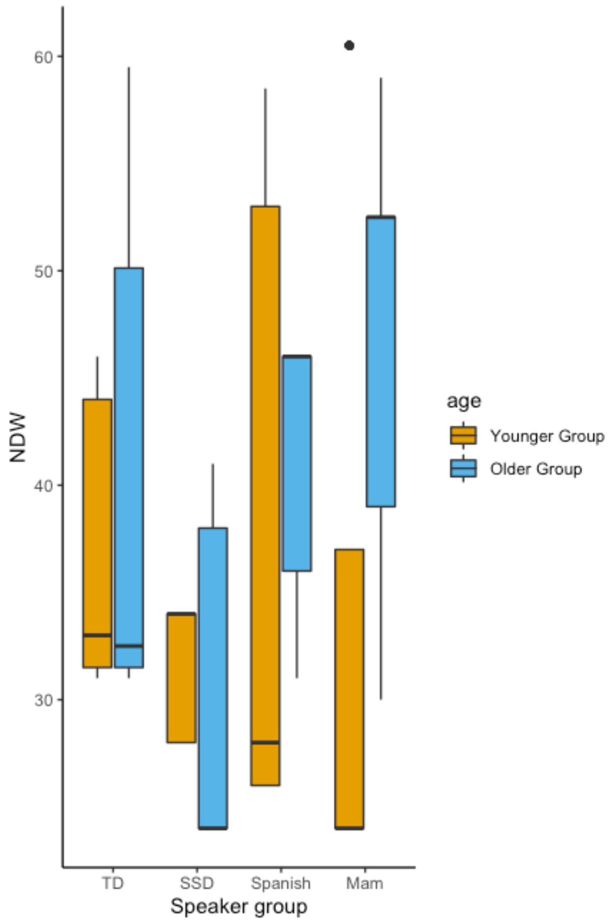


Figure 14. Number of different words (NDW) by age and speaker group.

Note. TD = monolingual English typically developing, SSD = monolingual English speech sound disorder, Spanish = typically developing Spanish-English bilingual, Mam = typically developing Mam-English bilingual.

6.2.2. Influence of Linguistic Ability on Expert GIAs

Table 16 displays the full model results of the linear mixed-effects model that examined the effect of linguistic ability on expert listener GIAs. There was a significant main effect of MLUm on GIA, indicating that the length and complexity of the children's utterances impacted intelligibility judgments. However, this was not the case for NDW, so the variety of vocabulary the children used did not have an impact on GIA.

Table 16. Linear mixed-effects model results with expert listener's GIA as the dependent variable.

	Model summary			
	β	SE	*t*	*p*
Fixed effects				
Intercept	79.60	5.19	15.34	<.001***
MLUm	1.88	.81	2.34	.026*
NDW	-.09	.10	-.86	.40
	s^2			
Random effects				
Speaker	14.99			
Batch	0.00			

Note. MLUm = mean length of utterance in morphemes, NDW = number of different words, ***$p < .001$, **$p < .01$, *$p < .05$.

Recall that in the Chapter 5 analyses of the speaker effects on GIA, only the effect of speaker group was significant; the effect of age group was not. So, in a second analysis, age group was entered as a control variable to remove the variance in expert GIA due to this factor and test whether language ability, as measured by MLUm and NDW, might explain the effect of speaker group on GIA. Results from this model are presented in Table 17, which shows that when age group is entered, MLUm is no longer a significant predictor. So, when the variance in MLUm and NDW due to age is removed by entering age group as its own factor in the analysis on GIA, MLUm is no longer significant predictor of GIA.

Table 17. Linear mixed-effects model results with expert listener's GIA as the dependent variable, adding in age group variable.

	Model summary			
	β	SE	t	p
Fixed effects				
Intercept	80.43	5.11	15.73	<.001***
MLUm	1.58	.83	1.91	.066
NDW	-.08	.10	-.78	.44
Age	2.36	1.94	1.22	.233
	s^2			
Random effects				
Speaker	13.78			
Batch	0.00			

Note. Reference for Age Group is younger group; MLUm = mean length of utterance in morphemes, NDW = number of different words, ***p < .001, **p < .01, *p < .05.

6.3. General Discussion

In this chapter, we examined the linguistic ability of the child speaker's and its influence on impressionistic intelligibility judgments (GIA). It was predicted that listeners would be influenced by children's language ability such that they would judge children's speech as more intelligible if the children produced longer, more complex utterances (MLUm) and more varied vocabulary (NDW). If the listeners were influenced by more advanced syntax and vocabulary rather than the speech signal, this would be problematic since they would be judging the wrong concept.

First, with regards to performance on the language variables, it was predicted that older children would have stronger language abilities compared to younger children. However, MLUm and NDW were not significant factors for age group or speaker group. The descriptive data do show these differences, but the inferential statistics do not, perhaps due to the small number of data points in each comparison group.

Next, we found MLUm to be a significant factor on GIA, but only in a reduced model that did not include age group or speaker group. When age was controlled for by adding it back in to the model, we see that MLUm loses its significance. Perhaps listeners were making allowances for language ability and adjusted their expectations of the children's level of understandability, as we postulated, they were doing with age. The result that MLUm and NDW were not significant predictors of GIA might suggest that these measures are not sensitive measures of language ability. However, there is ample evidence that these established measures are valid and reliable for monolingual (Charest et al., 2020; Fey et al., 2004; Moyle et al., 2011; Potratz et al, 2022; Rice et al., 2010) and bilingual school-age children (e.g., Bedore et al., 2006; Goldstein et al., 2010). On the other hand, perhaps expert listeners were in fact basing their judgments on speech factors (i.e., those that listeners used in the lab-based intelligibility task). This would have an implication for the Mam-English bilingual children since this speaker group was the one distinguished by GIA.

Recall that the Mam-English bilingual children had less exposure to English (*M* = 2.63 years, *SD* = 1.06) than the Spanish-English bilingual group (*M* = 4 years, *SD* = 1.93). Remote instruction due to the COVID-19 pandemic just prior to data collection likely contributed to the children's limited English proficiency. So, even though they had less exposure to English than the Spanish-English bilingual children, language ability still did not influence GIA. Less exposure to English could have negatively influenced the GIAs, since the speech fluency (i.e., smoothness of words and phrases being joined together) of Mam-English bilingual children was observed to be less than Spanish-English bilingual children. If it was the case that expert listeners were basing judgments on speech factors, then this suggests that an unfamiliar accent may indeed undermine GIA scores, even in expert listeners. As suggested in the background chapter, there may have been factors of racial bias at play that influenced listener's intelligibility judgments. Listeners may have expected the Mam-English bilingual children to sound more like the children with Spanish-influenced English based on their similar physical appearance. So, a mismatch occurs with the listener's expectations which then may have resulted in lower GIAs for the Mam-English bilingual group.

There were two surprising findings regarding the language ability measures in the bilingual groups given that these children were acquiring English as a second language. First, children in the Spanish-English bilingual group had the highest overall MLUm and NDW, out-performing the monolingual children. Second, children in the older bilingual Mam-English group had the highest NDW. Several studies have compared the performance of monolingual and bilingual children on these same syntactic and lexical development measures (e.g., Bedore et al., 2006; Gabani et al., 2009; Pearson et al., 1993). Most have found no differences, or that monolingual children have higher MLU and NDW compared to bilingual children. Our study is the first to find the opposite pattern. One explanation for this may be that the small number of child speakers in each group lent itself to influences of the unique individuals in each group. Additionally, it was observed that the many of the bilingual child speakers had a high level of comfort during the speaking tasks and took pride in the fact that they were recruited for their "special talent" of being bilingual. This may have led to their displaying their speaking skills during the language sample task in a way that the monolingual speakers did not. For example, the bilingual children seemed very interested in showing off their skills and used descriptive language to do so (e.g., "But the other kid saw the rabbit sticking out his ear under his backpack.").

To summarize, the overall findings of this chapter indicate that language ability, as we measured it with MLUm and NDW, did not influence GIA. Regarding the Mam-English bilingual children, even though they had less exposure to English than the Spanish-English bilingual children, language ability still does not influence GIA for them. So, there may be speech effects secondary to their having less exposure and/or there are factors of racial bias at play related to the mismatch with listener expectation based on their physical appearance (i.e., they appear to be Spanish-speakers, but do not speak with Spanish-influenced English).

CHAPTER VII

CONCLUSIONS AND FUTURE DIRECTIONS

Perceptual speech intelligibility judgments are critical to the evaluation of children for speech sound disorder (SSD) and are frequently the basis for clinical decision making. However, this type of assessment has its limitations. As Kent (1996) wrote, “knowing the nature of these failings and limitations is important in minimizing their undesired effects.” The current dissertation investigated some of these potential “failings and limitations” through an examination of the influences of speaker age, linguistic ability, and accent familiarity on speech-language pathologist’s speech intelligibility judgments. This chapter presents a summary of the main findings and the conclusions that were drawn regarding the analyses in the experimental chapters. In addition, directions for future research are suggested based on areas of need and the limitations of the current study. Clinical implications of the conclusions are also discussed.

7.1. The Usefulness of Global Intelligibility Assessments (GIA)

Our study found that GIA did not show differences between the monolingual English speech sound disorder (SSD) group and the typically developing monolingual English group, but the experimentally controlled lab-based measure of intelligibility (i.e., LIS) did show that these two groups were different in their speech. This calls into question the GIA results. Note that the parent ratings on the Intelligibility in Context Scale (ICS) also did not show this difference. Additionally, SLPs apparently make real accommodations in their assessments based on age and linguistic ability. This lack of sensitivity and objectivity of the measure may prompt one to ask why to continue eliciting spontaneous speech samples and GIAs. However, recall that some children will score in the average range on a single-word standardized test, but then still fail to be understood in spontaneous speech. So, an assessment of a child's connected speech is a vital component in a comprehensive evaluation.

Even with its flaws, GIA is still useful since it provides information that is not provided by single word tests, including what happens with connected speech (e.g., syllables and words strung together, prosodic features). But clearly, it is not a good measure by itself for identifying a speech sound disorder and it has already been shown to lack sensitivity to changes due to treatment (Lousada et al., 2014). So, despite the problems with GIA identified in this dissertation and those already established, GIA is providing important information that cannot be gained with another method, so it is worth keeping and using in combination with other measures. In other words, while these issues exist, an intelligibility judgment from a spontaneous speech sample is necessary to get the full picture of the child's functional communication.

Of note is that the children in the SSD group were only mildly impaired, producing errors that may have been considered developmental (e.g., gliding of /ɹ/) by the listeners. Recall that even their parents rated them about the same as their typically developing peers on the Intelligibility in Context Scale (ICS) (see Table 2). The mild severity of the children in the SSD group may have had an impact on GIA. The listeners did not have age information, so they may have viewed these children as having only developmental errors.

7.2. What Does a Global Intelligibility Assessment Measure?

There were several key findings from the current study that allow us to better identify what GIA is measuring and what it is sensitive to. We began our questioning with the basic assumption that there are top-down influences on the listener (e.g., listener's experience, exposure, and familiarity) when they are making intelligibility judgments. Additionally, we started with survey data that showed SLPs define intelligibility as the amount of speech that is understood. While the expert and lay listener groups were different in their global intelligibility assessments (GIA), there was a lack of interaction with either age group or speaker group with GIA. This indicated no impact of the expert listener's training and experience on identifying children of different ages or linguistic backgrounds with this measure. Less variance in expert listener's GIAs did show that SLPs have honed their skills relative to lay listeners, however not in a way that their judgments separated out age groups or speaker groups. So, we know that GIA is a subjective measure that was not impacted by expert listener's training.

Next, GIA was not sensitive to the age of the children in this study, whereas both the lab-based intelligibility score (LIS) and the comprehensibility rating (CR) did show an impact of the two age groups, with the older group (mean age of 7;8) receiving higher scores than the younger group (mean age of 6;1). So, the more objective measures did discern between age groups, but the more subjective measure did not. This shows a difference between the measures and reinforces the subjectivity of GIA. In this case however, it is positive that listeners compensated for their expectations based on age, so that was not a negative on the children. Recall that LIS was derived from audio files only, whereas GIA was determined from videos. Perhaps there was an impact being able to see how old the children appeared, and listeners were integrating what they knew about speech development and gave a pass to the younger children for their developmental errors.

Further characterizing GIA, we found it to be related to both intelligibility, as measured by LIS, and comprehensibility, as measured by CR. So, there appears to be an effect of speech features on GIA as well as a level of understandability. The effect of speech was shown with the correlation between GIA and LIS. Since these were positively correlated the assumption is that the purer measure of intelligibility, LIS, captured the elements of the speech signal that had an impact on intelligibility. Thus, we can assume there is a component of the GIA that does indeed address speech. Interestingly, the same was not true of our language measures.

When age group was controlled for, GIA was not sensitive to the language abilities of the child speakers. In other words, there was not an effect of language ability, as it was measured here with MLUm and NDW, on impressionistic intelligibility judgments. One possible explanation for this is that, just as with age, the listeners filtered out the language complexity of the child speakers and did indeed focus on the speech signal. This is another positive outcome since the goal of intelligibility assessments is on speech rather than language. So, we can conclude that expert listeners are in fact basing their judgments on speech factors, like that used in the lab-based intelligibility task. This suggests that an unfamiliar accent (i.e., that of the Mam-English bilingual children) may indeed undermine GIA, even in expert listeners. This is explored in detail in the next subsection.

7.3. The Effect of Mam-English Bilingual Speakers

The predictions regarding the Mam-English bilingual speaker group were that the speaker's accent, the listener's familiarity with the accent, and factors of racial bias would have an impact on their GIAs. We expected an impact of the visual images of the two groups of bilingual children, who have similar physical appearances but sound different from each other. Specifically, we anticipated that the speech of the Mam-English biligual children would be judged more harshly on the GIA than the speech of the Spanish-English bilingual children. There was indeed an effect of accent familiarity on GIA. Because the Mam-English bilingual children appeared as if they might be Spanish speakers (and this is borne out in their day-to-day lives as reported by their ELD teacher, K. Mitchell, personal communication, February 4, 2022), we posit that listener expectations were not met about what the children were anticipated to sound like based on their physical appearance. There was a mismatch between the listener's expectations and how the children sounded, resulting in lower intelligibility scores for the Mam-English bilingual children. This finding is certainly problematic given the frequent misdiagnoses of multilingual children as having speech and language impairments (Artiles et al., 2002; Sullivan, 2011).

7.4. Future Research Directions

The focus of this dissertation was on examining the influences of age, linguistic ability, and unfamiliar accent on global intelligibility assessments, with the goal of highlighting influential factors so that SLP training could improve the objectivity of intelligibility judgments. There are many other types of influences that are potential foci of future research. For example, since speech rate influences intelligibility in L2 learners

(Munro and Derwing, 2001), a future direction for research would be to examine the speech rate of child speakers to test for a correlation between rate and perceiver data on intelligibility. Further, Anderson-Hsieh and colleagues (1992) found stress, rhythm, intonation, and phrasing to have the strongest effect on nonnative speech in adults even over segmentals and syllable structure. Future research could explore if these findings would hold true in child speakers as well.

Research is also needed that more deeply examines SLP's intelligibility judgments of children with unfamiliar accents. For example, how much exposure to an unfamiliar accent is enough to remove the effect we found? And what tools might be useful to increase agreement of intelligibility judgments of speakers of an unfamiliar accent across listeners? Kent (1996) suggested using audiotaped and videotaped reference samples to assist in making auditory-perceptual assessments with greater listener agreement. At that time, reference samples were available for different vocal qualities and different classifications of dysarthria but had the potential for a wide range of communicative disorders (Kent, 1996). This idea is like when an adult client presents with accented speech and the clinician uses a website resource (e.g., *Speech Accent Archive*, n.d.) to analyze and compare accents. While there are no such known resources for child speech, there may be other opportunities for comparison, such as the speech of siblings or peers who speak the same languages. Along these lines, there is a need to examine other familiar and unfamiliar languages besides the ones that were analyzed in this dissertation.

7.5. Training and Clinical Implications

The findings from the current study offer several implications for training and clinical applications. Indeed, it was the aim of this study to identify ways to improve the objectivity of impressionistic judgments. So, what have we learned that can be applied in practice? One primary take-away is that SLPs can improve their perceptual ratings by consciously factoring in any lack of familiarity with an accent when assessing a bilingual child, acknowledging that there are listener-related factors that have an influence. These influences can be lessened by use of best practices when there is a mismatch in language between the client and clinician (McLeod et al, 2017), the correct utilization of phonemic inventory resources (e.g., American Speech-Language-Hearing Association, n.d.) and/or peer or sibling speech as a reference. Interpreters who speak the same dialect as the child can also be used as a resource. Referring to the Vietnamese child we gave as an example in the introduction chapter, this is a case in which the SLP would use such references to guide them in determining the eligibility of the child.

What additional steps can be taken to enhance training and hone the skills of SLPs with regard to making more objective intelligibility judgments? The explicit training on developmental speech sound and language changes with age that SLPs receive appears to have made an impact on their intelligibility judgments of children of different ages; they rated younger children with an allowance for age-expected differences. Therefore, SLPs should continue to be aware of developmental speech and language norms for age when they are making impressionistic intelligibility judgments. The training and exposure that they receive is vital to this. Memorizing, internalizing or simply referring to developmental norms is part of this process. The veteran SLP, who is also perhaps exposed to children of various developmental stages as a parent, uncle, or neighbor will be aware and familiar with errors that are developmentally appropriate and take that into consideration when assigning a percentage of understandability to the child they are assessing.

Finally, what can we conclude about the use of rating scales instead of orthographic transcription tasks to save time in a clinical setting? While some research has shown the validity of replacing transcription tasks with rating scales, more research is needed in this area, along with more nuanced suggestions that incorporate type of disorder and severity. For example, it has been recommended that a rating scale measure be used for more linguistically advanced children and that word identification tests be used for children with lower levels of language (Porter & Bradley, 1985). More of these types of recommendations are needed to further enhance the objectivity of intelligibility judgments.

APPENDIX A: BILINGUALISM QUESTIONAIRRE

1. List **all** the languages your child has been exposed to: ___________________
2. When your child was born, what language did they hear? __________________________
3. What dialect/region is native language from? ___________________
4. At what age was your child exposed to English? __________
5. Describe your child's schedule on a typical day to account for all linguistic interactions (such as mealtime, school, TV, entertainment). Use as many rows as needed.

Activity	Who is involved?	Weekend or Weekday?	English		Spanish or Mam (circle one)	
			Input	Output	Input	Output
Example. *School*	*Teacher, peers*	*weekday*	*x*	y	*0*	*0*
1.						
2.						
3.						
4.						
5.						
6.						
7.						
8.						
9.						
10.						
11.						
12.						
13.						
14.						
15.						
Percent	-------------		%	%	%	%

6. Proficiency in each language: check one rating for each language

Rating	English	Spanish or Mam (circle one)
4= child has native-like proficiency in this language		
3		
2		
1		
0=cannot speak this language at all		

APPENDIX B: MAM ARTICULATION SCREENER

English/Spanish	Mam: Todos Santos*	Transcription
1. red/rojo	**ky**aq	
2. tree/árbol	**tz**ee’	
3. two/dos	**k**aaba	
4. handkerchief/pańuelo	**s**u’tj	
5. horse/caballo	**tsh**eej	
6. lizard/lagartija	x**h**ule	
7. woman/mujer	mo**x**	
8. monkey/mono	**sh**maash	
9. throat/garganta	**ch**un	
10. lime/cal	**tx**un	
11. tomato/miltomate	**tsh’**u’sh	
12. dirty/sucio	**tz’**iil	
13. pointed/puntiagudo	**ch’**u’p	
14. Earth/tierra	**tx’**otx’	
15. home/casa	**h**aa’	
16. belt/cinturón	**p**ash	
17. bone/hueso	**b**aq	
18. soup/sopa	**t**a’l	
19. spark/chispa	**t’**ilpah	

*target sound is in bold

APPENDIX C: QUESTION PROMPTS FOR EACH STORY

Questions for *Dog Comes Home*

1. What do you think the girl is thinking here?
2. Why do you think she's putting the dog in her bag?
3. Why do you think the girl's getting so dirty?
4. Why is there a white dog in the bathtub now?
5. What do you think the mother's going to do now?
6. What would you say to the mom if you were the girl here?

Questions for *Bunny Goes to School*

1. Why do you think the bunny jumped out of the backpack?
2. Why do you think some students are afraid and some students are laughing?
3. What would you do if a bunny came to your school?
4. What was the boy's idea?
5. How did the mom know to come to the school?
6. Why do you think the mom came to the school?
7. What do you think will happen when the boy when he goes home?
8. What do you think the teacher's thinking now?

Questions for *Baseball Troubles*

1. What is the teacher thinking here?
2. What are the two boys doing with their baseball mits? Why?
3. Why doesn't the teacher know that the two boys broke the window?
4. What is the teacher saying to the girl with the headphones?
5. What do you think the girls who were playing jump rope are going to do? Why?
6. What should the girls do?
7. What would you do if you were the girl being blamed?
8. Have you ever been blamed for something you didn't do? Tell me what happened.

Questions for *Baseball Troubles*

1. Will the big boys share their ball with the little boys? How do you know?
2. What are the little boys thinking here? (Look at their expressions)
3. Why don't the big boys see that the little boys are playing with the ball?
4. What are the big boys thinking now?
5. What do you think the big boys will do if the ball falls out of the shirt?
6. What would you say to try to get out of trouble if you were the little boys?
7. Have you ever gotten into big trouble like this? What happed? How did you get out of it?

APPENDIX D: SAMPLE NARRATIVES

Narrative by a typically developing, monolingual English participant in the younger group for *Dog Comes Home*

E tell me the whole story?
#C alright.
C Dog.
C She found a dog under the porch.
C it was black.
C and the dog lick/ed her.
C and then it lick/ed her.
C and she was dirty.
C And then her mom said no dog/s allow/ed.
C And it was happy the dog.
C and she said %SHH to the dog and put it in her purse.
C And then the mom was surprise/ed to see her dirty.
C And then she sent her straight into the bath.
C and then she was surprise/ed to see a white dog in the bath.
C And the dirty bubble/s and some dog print/s.
C so she was like what?
C cause (there's a) there/'s a exclamation point and a question mark so she/'s like what?

Narrative by a typically developing, monolingual English participant in the older group for *Dog Comes Home*

E can you tell me the whole story?
C the girl was walk/ing around.
C and then she saw a dog under her porch.
C And then the dog came out.
C and she play/ed with the dog.
C And then the dog got her really dirty.
C And then she thought that her mom would say no_no_no doggie/s allow/ed in the house.
C And then she was say/ing SH to the dog.
C so she would/n't get caught.
C and put the dog in the pack.
C then she close/ed the pack.
C And when she got in the house her mom was like AH you get in the bath right now.
C And then the dog was wag/ing his tail out the back of the pack.
C And then the mom said get in the bath.
C And then she got in the bath.
C and the dog all clean and white.
C And the bath sud/s were all dirty with dirt.
C (And there was paw/s all over the) and there was paw/s on the bathtub.
C and there was paw/s on the he carpet and the floor.

C And the mom was like in her head she was like what the?
C What in the world happen/ed here?
C And then that/'s the end.

Narrative by a typically developing, Mam-English bilingual participant in the younger group for *Dog Comes Home*

C when the kid found a dog.
C It/'s (at) in her house.
C (It's so) it/'s wiggly (and it/'s so).
C When she saw the dog it was scared and nobody stealed|steal[EO:stole] him.
C and she grab him.
C (And he try) and then he lick her face.
C (And then she said) and then she laugh/ed.
C She think when her mom said no dog/s in our house.
C She think that.
C and she said be quiet I/'m go/ing to put you in my backpack.
C I mean my bag I mean.
C And then (then when she come to school) when she come to school (she was) her mom said why are dirty?
C X And she said I don't know.
C and she said that.
C And she said (go in the) go shower now.
C And she said okay.
C And then she go.
C (And then) and then (her mom when she come) when the mom come in the bathroom the dog it was there too.

Narrative by a typically developing, Spanish-English bilingual participant in the younger group for *Dog Comes Home*

C (um well) the girl was walk/ing all the way back home until she stop/ed because she saw a puppy down the big house.
C And the puppy came in.
C and then the puppy was so dirty it start/ed rub/ing all of that dirty (to her dau) to the girl.
C And then the puppy was start/ing lick/ing.
C And then the girl want/ed to take the puppy home.
C But then she remember/ed (no) no dog/s here missy.
C so then she was quiet I want to take you with me.
C so then she put the dog in the backpack.
C and then she went home.
C and then when she came her mom got surprise/ed she was all dirty.

C (she said) she said sweetheart you must take a shower.
C No!
C and so then she went to the X tub with her bag.
C And then she splat/ing with the dog.
C but the dog did/n't let something.
C but let the X.
C And then the mom came sweetheart how did this dog get in here?
C And the sweetheart was uh.
C and then her mom said I/'m mad and disappoint/ed.
C you/'re ground/ed.

APPENDIX E: LOW-PREDICTABILITY SENTENCES

Adapted from Stelmachowicz et al., 2000

1. Quick books look bright.
2. Wide pens swim fast.
3. Cats get good boats.
4. Feet drink hot bread.
5. Clocks catch old cows.
6. Cups give fat ducks.
7. Green hands don't fall.
8. Blue chairs draw well.
9. Late milk drank shirts.
10. Sad cars want chills.
11. Strange nails taste dark.
12. Hard corn feels mean.
13. High bears move holes.
14. Guys tell loud meat.
15. Now straws need cheese.
16. Drums pour tall pets.
17. True kings keep new.
18. Blocks can't run sharp.
19. Please shine some clowns.
20. Dull socks wag off.
21. Snow smells more tough.
22. Most birds knock tea.
23. Four rates close warm.
24. Large food jumps loose.
25. Long kids stay back.
26. Big apes grab sun.
27. Slow bugs itch far.
28. Black frogs bring Mom.
29. Rich men might pop.
30. Great gum hurts three.
31. Smart shoes lend play.
32. Tin cake sleeps hard.
33. Leave planes cool fun.
34. Cold worms leave toys.
35. Jokes fly like doors.
36. Feed Dad down tooth.
37. Dog poked sweet trains.
38. Trees beat small rain.
39. Sisters turn big trees.
40. Clean bee wrecked friends.

REFERENCES CITED

Abbeduto, L. (2021, June). *Rethinking clinical trials in neurodevelopmental disabilities: Contributions from language assessment and intervention* [Invited Speaker]. Symposium on Research in Child Language Disorders, Madison, WI. https://srcldconference.com/the-2021-conference/

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 520–529. https://doi.org/10.1037/a0013552

Aissen, J., England, N. C., & Zavala, R. (2017). *The Mayan languages*. Taylor & Francis.

Allison, K., & Hustad, K. C. (2014). Impact of sentence length and phonetic complexity on intelligibility of 5-year-old children with cerebral palsy. *International Journal of Speech-Language Pathology*, *16*(4), 396–407. https://doi.org/DOI: 10.3109/17549507.2013.876667

Allison, K. M., Russell, M., & Hustad, K. C. (2021). Reliability of perceptual judgments of phonetic accuracy and hypernasality among speech-language pathologists for children with dysarthria. *American Journal of Speech-Language Pathology*, *30*(3S), 1558–1571. https://doi.org/10.1044/2020_AJSLP-20-00144

American Speech-Language-Hearing Association. (n.d.). *Phonemic inventories and cultural and linguistic information across languages*. Retrieved November 12, 2021, from https://www.asha.org/practice/multicultural/Phono/

American Speech-Language-Hearing Association. (2022). 2021 Demographic profile of ASHA members providing multilingual services. Available from www.asha.org.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, *42*(4), 529–555. https://doi.org/10.1111/j.1467-1770.1992.tb01043.x

Andrews, N., & Fey, M. E. (1986). Analysis of the speech of phonologically impaired children in two sampling conditions. *Language, Speech, and Hearing Services in Schools*, *17*(3), 187–198. https://doi.org/10.1044/0161-1461.1703.187

Arias, G., & Friberg, J. (2017). Bilingual language assessment: Contemporary versus recommended practice in American schools. *Language, Speech, and Hearing Services in Schools*, *48*(1), 1–15. https://doi.org/10.1044/2016_LSHSS-15-0090

Artiles, A. J., Harry, B., Reschly, D. J., & Chinn, P. C. (2002). Over-identification of students of color in special education: A critical overview. *Multicultural Perspectives*, *4*(1), 3–10. https://doi.org/10.1207/S15327892MCP0401_2

Audacity Team. (2012). *Audacity* (3.1.3) [Computer software]. www.audacityteam.org

Ball, M. J. (2012). Vowels and consonants of the world's languages. In S. McLeod & B.A. Goldstein (Eds.), *Multilingual aspects of speech sound disorders in children* (pp. 32–41). Multilingual Matters.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Baudonck, N. L. H., Buekers, R., Gillebert, S., & Van Lierde, K. M. (2009). Speech intelligibility of Flemish children as judged by their parents. *Folia Phoniatrica et Logopaedica*, *61*(5), 288–295. https://doi.org/10.1159/000235994

Bedore, L. M., Fiestas, C. E., Peña, E. D., & Nagy, V. J. (2006). Cross-language comparisons of maze use in Spanish and English in functionally monolingual and bilingual children. *Bilingualism: Language and Cognition*, *9*(3), 233–247. https://doi.org/10.1017/S1366728906002604

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America*, *114*(3), 1600–1610. https://doi.org/10.1121/1.1603234

Bent, T., Bradlow, A. R., & Smith, B. L. (2007). Segmental errors in different word positions and their effects on intelligibility of non-native speech. In O.-S. Bohn and M.J. Munro (Eds.) *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 331–347). J. Benjamins Pub.

Bent, T., & Frush Holt, R. (2013). The influence of talker and foreign-accent variability on spoken word identification. *The Journal of the Acoustical Society of America*, *133*(3), 1677–1686. https://doi.org/10.1121/1.4776212

Bernthal, J. E., Bankson, N. W., & Flipsen, P. (2022). *Speech sound disorders in children: Articulation & phonological disorders* (Ninth edition). Paul H. Brookes Publishing Co., Inc.

Bilingual Speechie. (n.d.). *Pre-K and Kindergarten Artic and Language Screener Spanish*. https://bilingualspeechie.com/collections/screeners/products/preschool-kindergarten-language-articulation-screener-spanish

Bleile, K. (2002). Evaluating articulation and phonological disorders when the clock Is running. *American Journal of Speech-Language Pathology*, *11*(3), 243–249. https://doi.org/10.1044/1058-0360(2002/026)

Bleile, K. M. (2020). *Speech sound disorders: For class and clinic* (Fourth edition). Plural Publishing Inc.

Boersma, P., & Weenink, D. (2022). *Praat: Doing phonetics by computer* (Version 6.1.27) [Computer software]. http://www.praat.org.

Borrie, S. A., McAuliffe, M. J., Liss, J. M., Kirk, C., O'Beirne, G. A., & Anderson, T. (2012). Familiarisation conditions and the mechanisms that underlie improved recognition of dysarthric speech. *Language and Cognitive Processes*, *27*(7–8), 1039–1055. https://doi.org/10.1080/01690965.2011.610596

Brown, V. A., McLaughlin, D. J., Strand, J. F., & Van Engen, K. J. (2020). Rapid adaptation to fully intelligible nonnative-accented speech reduces listening effort. *Quarterly Journal of Experimental Psychology*, *73*(9), 1431–1443. https://doi.org/10.1177/1747021820916726

Brown, V. A., & Strand, J. F. (2019). About face: Seeing the talker improves spoken word recognition but increases listening effort. *Journal of Cognition*, *2*(1), 44. https://doi.org/10.5334/joc.89

Brunnegård, K., Lohmander, A., & van Doorn, J. (2009). Untrained listeners' ratings of speech disorders in a group with cleft palate: A comparison with speech and language pathologists, ratings. *International Journal of Language & Communication Disorders*, *44*(5), 656–674. https://doi.org/10.1080/13682820802295203

Byers-Heinlein, K., Esposito, A. G., Winsler, A., Marian, V., Castro, D. C., & Luk, G. (2019). The case for measuring and reporting bilingualism in developmental research. *Collabra: Psychology*, *5*(1), 37. https://doi.org/10.1525/collabra.233

Carrow-Woolfolk, E. (2017). *Comprehensive Assessment of Spoken Language-Second Edition* (2nd ed.). Pearson Assessments.

Charest, M., Skoczylas, M. J., & Schneider, P. (2020). Properties of lexical diversity in the narratives of children with typical language development and developmental language disorder. *American Journal of Speech-Language Pathology*, 1–17. https://doi.org/10.1044/2020_AJSLP-19-00176

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). L. Erlbaum Associates.

Coplan, J., & Gleason, J. R. (1988). Unclear speech: Recognition and significance of unintelligible speech in preschool children. *Pediatrics*, *82*(3), 447–452.

Counts, J., Katsiyannis, A., & Whitford, D. K. (2018). Culturally and linguistically diverse learners in special education: English learners. *NASSP Bulletin*, *102*(1), 5–21. https://doi.org/10.1177/0192636518755945

Crowley, C. & Baigorri, M. (2015). *School-age language assessment measures*. LEADERSproject. https://www.leadersproject.org/disability-evaluation/school-age-language-assessment-measures-slam/

Cycyk, L. M., De Anda, S., Ramsey, K. L., Sheppard, B. S., & Zuckerman, K. (2022). Moving through the pipeline: Ethnic and linguistic disparities in special education from birth through age five. *Educational Researcher*, 0013189X2211202. https://doi.org/10.3102/0013189X221120262

Darboe, K. (2003). New immigrants in Minnesota: The Somali immigration and assimilation. *Journal of Developing Societies*, *19*(4), 458–472. https://doi.org/10.1177/0169796X0301900402

Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology*, *19*(2), 294–304. https://doi.org/10.1037/0893-3200.19.2.294

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, *20*, 1–16. http://www.jstor.org/stable/44488664

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, *42*(4), 476–490. https://doi.org/10.1017/S026144480800551X

Eadie, P., Morgan, A., Ukoumunne, O. C., Ttofari Eecen, K., Wake, M., & Reilly, S. (2015). Speech sound disorder at 4 years: Prevalence, comorbidities, and predictors in a community cohort of children. *Developmental Medicine & Child Neurology*, *57*(6), 578–584. https://doi.org/10.1111/dmcn.12635

Eakins, B. W. (1969). Research in slow-played speech: Listener training, consonant errors, and sex and age differences in speaker intelligibility. *Central States Speech Journal*, *20*(4), 302–307. https://doi.org/10.1080/10510976909362982

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research*, *47*(2), 421–436. https://doi.org/10.1044/1092-4388(2004/034)

Eisenberg, S. L., & Hitchcock, E. R. (2010). Using standardized tests to inventory consonant and vowel production: A comparison of 11 tests of articulation and phonology. *Language, Speech, and Hearing Services in Schools*, *41*(4), 488–503. https://doi.org/10.1044/0161-1461(2009/08-0125)

Ellis, L. W., & Beltyukova, S. A. (2008). Effects of training on naïve listeners' judgments of the speech intelligibility of children with severe-to-profound hearing loss. *Journal of Speech, Language, and Hearing Research*, *51*(5), 1114–1123. https://doi.org/10.1044/1092-4388(2008/06-0217)

England, N. C. (2011). *A grammar of Mam, a Mayan Language*. University of Texas Press.

England, N. C. (2017). Mam. In J. Aissen, N.C. England, & R.Z. Maldonado (Eds.) *The Mayan languages* (pp. 500–532). Taylor & Francis.

Ertmer, D. J. (2010). Relationships between speech intelligibility and word articulation scores in children with hearing loss. *Journal of Speech Language and Hearing Research*, *53*(5), 1075–1086. https://doi.org/10.1044/1092-4388(2010/09-0250)

Evans, K. E., Munson, B., & Edwards, J. (2018). Does speaker race affect the assessment of children's speech accuracy? A comparison of speech-language pathologists and clinically untrained listeners. *Language, Speech, and Hearing Services in Schools*, *49*(4), 906–921. https://doi.org/10.1044/2018_LSHSS-17-0120

Fabiano-Smith, L. (2019). Standardized tests and the diagnosis of speech sound disorders. *Perspectives of the ASHA Special Interest Groups*, *4*(1), 58–66. https://doi.org/10.1044/2018_PERS-SIG1-2018-0018

Fabiano-Smith, L., & Barlow, J. A. (2010). Interaction in bilingual phonological acquisition: Evidence from phonetic inventories. *International Journal of Bilingual Education and Bilingualism*, *13*(1), 81–97. https://doi.org/10.1080/13670050902783528

Fabiano-Smith, L., & Goldstein, B. A. (2010). Phonological acquisition in bilingual Spanish–English speaking children. *Journal of Speech, Language, and Hearing Research*, *53*(1), 160–178. https://doi.org/10.1044/1092-4388(2009/07-0064)

Fabiano-Smith, L., Privette, C., & An, L. (2021). Phonological measures for bilingual Spanish-English–speaking preschoolers: The language combination effect. *Journal of Speech, Language, and Hearing Research*, *64*(10), 3942–3968. https://doi.org/10.1044/2021_JSLHR-21-00008

Fey, M. E., Catts, H. W., Proctor-Williams, K., Tomblin, J. B., & Zhang, X. (2004). Oral and written story composition skills of children with language impairment. *Journal of Speech, Language, and Hearing Research*, *47*(6), 1301–1318. https://doi.org/10.1044/1092-4388(2004/098)

Flipsen, P. (2006). Measuring the intelligibility of conversational speech in children. *Clinical Linguistics & Phonetics*, *20*(4), 303–312. https://doi.org/10.1080/02699200400024863

Frizelle, P., Thompson, P. A., Mcdonald, D., & Bishop, D. V. M. (2018). Growth in syntactic complexity between four years and adulthood: Evidence from a narrative task. *Journal of Child Language*, *45*(5), 1174–1197. https://doi.org/10.1017/S0305000918000144

Fuse, A., Navichkova, Y., & Alloggio, K. (2018). Perception of intelligibility and qualities of non-native accented speakers. *Journal of Communication Disorders*, *71*, 37–51. https://doi.org/10.1016/j.jcomdis.2017.12.006

Gabani, K., Sherman, T., Solorio, T., Liu, Y., Bedore, L., & Peña, E. (2009). A corpus-based approach for the prediction of language impairment in monolingual English and Spanish-English bilingual children. In *In Proceedings of human language technologies: The 2009 annual conference of the North American chapter of the Association for Computational Linguistics* (pp. 46–55).

Gass, S., & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning*, *34*(1), 65–87. https://doi.org/10.1111/j.1467-1770.1984.tb00996.x

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(1), 84–95. https://doi.org/10.1037/0278-7393.25.1.84

Gildersleeve-Neumann, C. E., Kester, E. S., Davis, B. L., & Peña, E. D. (2008). English speech sound development in preschool-aged children from bilingual English–Spanish environments. *Language, Speech, and Hearing Services in Schools*, *39*(3), 314–328. https://doi.org/10.1044/0161-1461(2008/030)

Goehl, H., & Martin, P. (1987, November). *Intelligibility and familiarity: Is experience enough*. Annual Convention of the American Speech-Language-Hearing Association, New Orleans, LA.

Goldman, R., & Fristoe, M. (2015). *Goldman-Fristoe Test of Articulation-Third Edition* (GFTA-3). American Guidance Service.

Goldstein, B. (2000). *Resource guide on cultural and linguistic diversity*. Singular Pub. Group.

Goldstein, B. A., & Bunta, F. (2012). Positive and negative transfer in the phonological systems of bilingual speakers. *International Journal of Bilingualism*, *16*(4), 388–401. https://doi.org/10.1177/1367006911425817

Goldstein, B. A., Bunta, F., Lange, J., Rodriguez, J., & Burrows, L. (2010). The effects of measures of language experience and language ability on segmental accuracy in bilingual children. *American Journal of Speech-Language Pathology*, *19*(3), 238–247. https://doi.org/10.1044/1058-0360(2010/08-0086)

Goldstein, B. A., Fabiano, L., & Washington, P. S. (2005). Phonological skills in predominantly English-speaking, predominantly Spanish-speaking, and Spanish-English bilingual children. *Language, Speech, and Hearing Services in Schools*, *36*(3), 201–218. https://doi.org/10.1044/0161-1461(2005/021)

Gordon-Brannan, M., & Hodson, B. W. (2000). Intelligibility/severity measurements of prekindergarten children's speech. *American Journal of Speech-Language Pathology*, *9*(2), 141–150. https://doi.org/10.1044/1058-0360.0902.141

Grigos, M. I. (2016). Speech sound disorders: What's motor got to do with it? *Perspectives of the ASHA Special Interest Groups*, *1*(1), 75–87. https://doi.org/10.1044/persp1.SIG1.75

Guo, L.-Y., & Eisenberg, S. (2015). Sample length affects the reliability of language sample measures in 3-year-olds: Evidence from parent-elicited conversational samples. *Language, Speech, and Hearing Services in Schools*, *46*(2), 141–153. https://doi.org/10.1044/2015_LSHSS-14-0052

Hammer, C. S., Komaroff, E., Rodriguez, B. L., Lopez, L. M., Scarpino, S. E., & Goldstein, B. (2012). Predicting Spanish–English bilingual children's language abilities. *Journal of Speech, Language, and Hearing Research*, *55*(5), 1251–1264. https://doi.org/10.1044/1092-4388(2012/11-0016)

Hansen, K., & Dovidio, J. F. (2016). Social dominance orientation, nonnative accents, and hiring recommendations. *Cultural Diversity and Ethnic Minority Psychology*, *22*(4), 544–551. https://doi.org/10.1037/cdp0000101

Hashemi Hosseinabad, H., Ishikawa, K., & Washington, K. (2021). Agreements between speech language pathologists and naïve listeners' judgements of Intelligibility in children with cleft palate. *Clinical Linguistics & Phonetics*, 1–19. https://doi.org/10.1080/02699206.2021.1983021

Hasson, N., Camilleri, B., Jones, C., Smith, J., & Dodd, B. (2013). Discriminating disorder from difference using dynamic assessment with bilingual children. *Child Language Teaching and Therapy*, *29*(1), 57–75. https://doi.org/10.1177/0265659012459526

Howard, S. J., & Heselwood, B. C. (2002). Learning and teaching phonetic transcription for clinical purposes. *Clinical Linguistics & Phonetics*, *16*(5), 371–401. https://doi.org/10.1080/02699200210135893

Huang, B., Alegre, A., & Eisenberg, A. (2016). A cross-linguistic investigation of the effect of raters' accent familiarity on speaking assessment. *Language Assessment Quarterly*, *13*(1), 25–41. https://doi.org/10.1080/15434303.2015.1134540

Hustad, K. C. (2006). Estimating the intelligibility of speakers with dysarthria. *Folia Phoniatrica et Logopaedica*, *58*(3), 217–228. https://doi.org/10.1159/000091735

Hustad, K. C. (2008). The relationship between listener comprehension and intelligibility scores for speakers with dysarthria. *Journal of Speech, Language, and Hearing Research*, *51*(3), 562–573. https://doi.org/10.1044/1092-4388(2008/040)

Hustad, K. C. (2012). Speech intelligibility in children with speech disorders. *Perspectives on Language Learning and Education*, *19*(1), 7–11. https://doi.org/10.1044/lle19.1.7

Hustad, K. C., & Cahill, M. A. (2003). Effects of presentation mode and repeated familiarization on intelligibility of dysarthric speech. *American Journal of Speech-Language Pathology*, *12*(2), 198–208. https://doi.org/10.1044/1058-0360(2003/066)

Hustad, K. C., Mahr, T., Natzke, P. E. M., & Rathouz, P. J. (2020). Development of speech intelligibility between 30 and 47 months in typically developing children: A cross-sectional study of growth. *Journal of Speech, Language, and Hearing Research*, *63*(6), 1675–1687. https://doi.org/10.1044/2020_JSLHR-20-00008

Hustad, K. C., Mahr, T. J., Natzke, P., & Rathouz, P. J. (2021). Speech development between 30 and 119 months in typical children I: Intelligibility growth curves for single-word and multiword productions. *Journal of Speech, Language, and Hearing Research*, *64*(10), 3707–3719. https://doi.org/10.1044/2021_JSLHR-21-00142

Hustad, K. C., Oakes, A., & Allison, K. (2015). Variability and diagnostic accuracy of speech intelligibility scores in children. *Journal of Speech, Language, and Hearing Research*, *58*(6), 1695–1707. https://doi.org/10.1044/2015_JSLHR-S-14-0365

Hustad, K. C., Schueler, B., Schultz, L., & DuHadway, C. (2012). Intelligibility of 4-year-old children with and without cerebral palsy. *Journal of Speech, Language, and Hearing Research*, *55*(4), 1177–1189. https://doi.org/10.1044/1092-4388(2011/11-0083)

Iacono, T. A. (1998). Analysis of the phonological skills of children with Down syndrome from single word and connected speech samples. *International Journal of Disability, Development and Education*, *45*(1), 57–73. https://doi.org/10.1080/1034912980450105

Ishikawa, K., Webster, J., & Ketring, C. (2021). Agreement between transcription- and rating-based intelligibility measurements for evaluation of dysphonic speech in noise. *Clinical Linguistics & Phonetics*, *35*(10), 983–995. https://doi.org/10.1080/02699206.2020.1852602

Jasso, J., & Potratz, J. R. (2020). Assessing speech sound disorders in school-age children from diverse language backgrounds: A tutorial with three case studies. *Perspectives of the ASHA Special Interest Groups*, *5*(3), 714–725. https://doi.org/10.1044/2020_PERSP-19-00151

Jing, L., & Grigos, M. I. (2022). Speech-language pathologists' ratings of speech accuracy in children with speech sound disorders. *American Journal of Speech-Language Pathology*, *31*(1), 419–430. https://doi.org/10.1044/2021_AJSLP-20-00381

Katz, L. A., Maag, A., Fallon, K. A., Blenkarn, K., & Smith, M. K. (2010). What makes a caseload (un)manageable? School-based speech-language pathologists speak. *Language, Speech, and Hearing Services in Schools*, *41*(2), 139–151. https://doi.org/10.1044/0161-1461(2009/08-0090)

Kent, R. D. (1996). Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology*, *5*(3), 7–23. https://doi.org/10.1044/1058-0360.0503.07

Kent, R., Weismer, G., Kent, J., & Rosenbek, J. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders*, *54*, 482–499. https://doi.org/10.1044/jshd.5404.482

Klein, H. B., Grigos, M. I., McAllister Byun, T., & Davidson, L. (2012). The relationship between inexperienced listeners' perceptions and acoustic correlates of children's /r/ productions. *Clinical Linguistics & Phonetics*, *26*(7), 628–645. https://doi.org/10.3109/02699206.2012.682695

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203. https://doi.org/10.1037/a0038695

Kutlu, E., Tiv, M., Wulff, S., & Titone, D. (2022). Does race impact speech perception? An account of accented speech in two different multilingual locales. *Cognitive Research: Principles and Implications*, *7*(1), 7. https://doi.org/10.1186/s41235-022-00354-0

Kwiatkowski, J., & Shriberg, L. D. (1992). Intelligibility assessment in developmental phonological disorders: Accuracy of caregiver gloss. *Journal of Speech, Language, and Hearing Research*, *35*(5), 1095–1104. https://doi.org/10.1044/jshr.3505.1095

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000). Prevalence and natural history of primary speech and language delay: Findings from a systematic review of the literature. *International Journal of Language and Communication Disorders*, *35*(2), 165–188.

Levy, E. S., Chang, Y. M., Ancelle, J. A., & McAuliffe, M. J. (2017). Acoustic and perceptual consequences of speech cues for children with dysarthria. *Journal of Speech Language and Hearing Research*, *60*(6S), 1766–1779. https://doi.org/10.1044/2017_JSLHR-S-16-0274

Lewis, B. A., Freebairn, L., Tag, J., Ciesla, A. A., Iyengar, S. K., Stein, C. M., & Taylor, H. G. (2015). Adolescent outcomes of children with early speech sound disorders with and without language impairment. *American Journal of Speech-Language Pathology*, *24*(2), 150–163. https://doi.org/10.1044/2014_AJSLP-14-0075

Loban, W. (1976). *Language development: Kindergarten through grade twelve*. National Council of Teachers of English.

Lohmander, A., Klintö, K., Schalling, E., Szabo Portela, A., Johansson, K., & McAllister, A. (2021). Students take charge of Learning – Using e-learning in Perceptual Assessment in Speech–Language Pathology. *Scandinavian Journal of Educational Research*, *65*(3), 468–480. https://doi.org/10.1080/00313831.2020.1716064

Lousada, M., Jesus, L. M. T., Hall, A., & Joffe, V. (2014). Intelligibility as a clinical outcome measure following intervention with children with phonologically based speech-sound disorders: Intelligibility as a clinical outcome measure. *International Journal of Language & Communication Disorders*, *49*(5), 584–601. https://doi.org/10.1111/1460-6984.12095

Lüdecke, D. (2021). *sjstats: Statistical Functions for Regression Models (Version 0.18.1)*. https://CRAN.R-project.org/package=sjstats

Ma, W. J., Zhou, X., Ross, L. A., Foxe, J. J., & Parra, L. C. (2009). Lip-reading aids word recognition most in moderate noise: A Bayesian explanation using high-dimensional feature space. *PLoS ONE*, *4*(3), e4638. https://doi.org/10.1371/journal.pone.0004638

MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*, *21*, 131–141.

McFarlane, S. C., Holt-Romeo, T. L., Lavorato, A. S., & Warner, L. (1991). Unilateral vocal fold paralysis: Perceived vocal quality following three methods of treatment. *American Journal of Speech-Language Pathology*, *1*(1), 45–48. https://doi.org/10.1044/1058-0360.0101.45

McLaughlin, D. J., Baese-Berk, M. M., Bent, T., Borrie, S. A., & Van Engen, K. J. (2018). Coping with adversity: Individual differences in the perception of noisy and accented speech. *Attention, Perception, & Psychophysics*, *80*(6), 1559–1570. https://doi.org/10.3758/s13414-018-1537-4

McLaughlin, D. J., & Van Engen, K. J. (2020). Task-evoked pupil response for accurately recognized accented speech. *The Journal of the Acoustical Society of America*, *147*(2), EL151–EL156. https://doi.org/10.1121/10.0000718

McLeod, S. (2012a). Multilingual speech assessment. In S. McLeod & B. Goldstein (Eds), *Multilingual aspects of speech sounds disorders in children* (pp. 113–143). Multilingual Matters.

McLeod, S. (2012b). Translation to practice: Creating sampling tools to assess multilingual children's speech. In S. McLeod & B. Goldstein (Eds), *Multilingual aspects of speech sounds disorders in children* (pp. 144–153). Multilingual Matters.

McLeod, S., Crowe, K., & Shahaeian, A. (2015). Intelligibility in context scale: Normative and validation data for English-speaking preschoolers. *Language Speech and Hearing Services in Schools*, *46*(3), 266. https://doi.org/10.1044/2015_LSHSS-14-0120

McLeod, S., Harrison, L. J., McAllister, L., & McCormack, J. (2013). Speech sound disorders in a community study of preschool children. *American Journal of Speech-Language Pathology*, *22*(3), 503–522. https://doi.org/10.1044/1058-0360(2012/11-0123)

McLeod, S., Harrison, L. J., & McCormack, J. (2012). The Intelligibility in Context Scale: Validity and reliability of a subjective rating measure. *Journal of Speech Language and Hearing Research*, *55*(2), 648–656. https://doi.org/10.1044/1092-4388(2011/10-0130)

McLeod, S., Verdon, S., Baker, E., Ball, M. J., Ballard, E., David, A. B., Bernhardt, B. M., Bérubé, D., Blumenthal, M., Bowen, C., Brosseau-Lapré, F., Bunta, F., Crowe, K., Cruz-Ferreira, M., Davis, B., Fox-Boyer, A., Gildersleeve-Neumann, C., Grech, H., Goldstein, B., … Zharkova, N. (2017). Tutorial: Speech assessment for multilingual children who do not speak the same language(s) as the speech-language pathologist. *American Journal of Speech-Language Pathology*, *26*(3), 691–708. https://doi.org/10.1044/2017_AJSLP-15-0161

Mahr, T. J., Rathouz, P. J., & Hustad, K. C. (2020). Longitudinal growth in intelligibility of connected speech from 2 to 8 years in children with cerebral palsy: A novel Bayesian approach. *Journal of Speech, Language, and Hearing Research*, *63*(9), 2880–2893. https://doi.org/10.1044/2020_JSLHR-20-00181

Marcus, A. P. (2020). Where Is the geography? The geographical predicaments of the panethnic terms Hispanic and Latino. *Journal of Latin American Geography*, *19*(2), 170–190. https://doi.org/10.1353/lag.2020.0026

Martin, N. A., & Brownell, R. (2011). *Expressive One-Word Picture Vocabulary Test—4* (EOWPVT-4). Academic Therapy Publications.

Martin, N. A., & Brownell, R. (2011). *Receptive One-Word Picture Vocabulary Test—4* (ROWPVT-4). Academic Therapy Publications.

Meir, N., Walters, J., & Armon-Lotem, S. (2016). Disentangling SLI and bilingualism using sentence repetition tasks: The impact of L1 and L2 properties. *International Journal of Bilingualism*, *20*(4), 421–452. https://doi.org/10.1177/1367006915609240

Melguy, Y. V., & Johnson, K. (2021). General adaptation to accented English: Speech intelligibility unaffected by perceived source of non-native accent. *The Journal of the Acoustical Society of America*, *149*(4), 2602–2614. https://doi.org/10.1121/10.0004240

Miller, J.F., Andriacchi, K., & Nockerts, A., (2019). *Assessing Language Production Using SALT Software: A Clinician's Guide to Language Sample Analysis* (3rd ed.). SALT Software, LLC.

Miller, J., & Iglesias, A. (2019). *Systematic Analysis of Language Transcripts (SALT), Student Version 20-1*. SALT Software, LLC.

Miller, N. (2013). Measuring up to speech intelligibility. *International Journal of Language & Communication Disorders*, *48*(6), 601–612. https://doi.org/10.1111/1460-6984.12061

Monsen, R. B. (1983). The oral speech intelligibility of hearing-impaired talkers. *Journal of Speech and Hearing Disorders*, *48*(3), 286–296. https://doi.org/10.1044/jshd.4803.286

Morrison, J. A., & Shriberg, L. D. (1992). Articulation testing versus conversational speech sampling. *Journal of Speech, Language, and Hearing Research*, *35*(2), 259–273. https://doi.org/10.1044/jshr.3502.259

Moyle, M. J., Karasinski, C., Ellis-Weismer, S., & Gorman, B. K. (2011). Grammatical morphology in school-age children with and without language impairment: A discriminant function analysis. *Language, Speech, and Hearing Services in Schools*, *42*(4), 550–560. https://doi.org/10.1044/0161-1461(2011/10-0029)

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*(1), 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Munro, M. J., & Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition*, *23*(4), 451–468. http://www.jstor.org/stable/44486957

Munro, M. J., Derwing, T. M., & Sato, K. (2006). Salient accents, covert attitudes: Consciousness-raising for pre-service second language teachers. *Prospect*, *21*(1), 67–79.

Munson, B., Edwards, J., Schellinger, S. K., Beckman, M. E., & Meyer, M. K. (2010). Deconstructing phonetic transcription: Covert contrast, perceptual bias, and an extraterrestrial view of *Vox Humana*. *Clinical Linguistics & Phonetics*, *24*(4–5), 245–260. https://doi.org/10.3109/02699200903532524

Munson, B., Johnson, J. M., & Edwards, J. (2012). The role of experience in the perception of phonetic detail in children's speech: A comparison between speech-language pathologists and clinically untrained listeners. *American Journal of Speech-Language Pathology*, *21*(2), 124–139. https://doi.org/10.1044/1058-0360(2011/11-0009)

Namasivayam, A. K., Pukonen, M., Goshulak, D., Yu, V. Y., Kadis, D. S., Kroll, R., Pang, E. W., & De Nil, L. F. (2013). Relationship between speech motor control and speech intelligibility in children with speech sound disorders. *Journal of Communication Disorders*, *46*(3), 264–280. https://doi.org/10.1016/j.jcomdis.2013.02.003

Nelson, M., & Wilson, L. (2021). Implicit bias and multilingual assessment in school-based speech-language pathologists. *Perspectives of the ASHA Special Interest Groups*, *6*(6), 1690–1704. https://doi.org/10.1044/2021_PERSP-21-00113

Nip, I. S. B., & Blumenfeld, H. K. (2015). Proficiency and linguistic complexity influence speech motor control and performance in Spanish language learners. *Journal of Speech, Language, and Hearing Research*, *58*(3), 653–668. https://doi.org/10.1044/2015_JSLHR-S-13-0299

Nippold, M. A. (2009). School-Age Children Talk About Chess: Does Knowledge Drive Syntactic Complexity? *Journal of Speech, Language, and Hearing Research*, *52*(4), 856–871. https://doi.org/10.1044/1092-4388(2009/08-0094)

Ockey, G. J., & French, R. (2016). From one to multiple accents on a test of L2 listening comprehension. *Applied Linguistics*, *37*(5), 693–715. https://doi.org/10.1093/applin/amu060

Panagos, J. M., & Prelock, P. A. (1982). Phonological constraints on the sentence productions of language-disordered children. *Journal of Speech, Language, and Hearing Research*, *25*(2), 171–177. https://doi.org/10.1044/jshr.2502.171

Park, M. S. (2020). Rater effects on L2 oral assessment: Focusing on accent familiarity of L2 teachers. *Language Assessment Quarterly*, *17*(3), 231–243. https://doi.org/10.1080/15434303.2020.1731752

Pascoe, M., Stackhouse, J., & Wells, B. (2006). *Persisting speech difficulties in children*. Wiley.

Paul, R., & Shriberg, L. D. (1982). Associations between phonology and syntax in speech-delayed children. *Journal of Speech, Language, and Hearing Research*, *25*(4), 536–547. https://doi.org/10.1044/jshr.2504.536

Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning*, *43*(1), 93–120.

Peppé, S. (2012). Prosody in the world's languages. In S. McLeod & B.A. Goldstein (Eds.), *Multilingual aspects of speech sound disorders in children* (pp. 42–52). Multilingual Matters.

Pollock, K. E. (1991). The identification of vowel errors using traditional articulation or phonological process test stimuli. *Language, Speech, and Hearing Services in Schools*, *22*(2), 39–50. https://doi.org/10.1044/0161-1461.2202.39

Porter, K., & Bradley, S. (1985). A comparison of three speech intelligibility measures for deaf students. *American Annals of the Deaf*, *130*(6), 514–525.

Potratz, J. R., Gildersleeve-Neumann, C., & Redford, M. A. (2022). Measurement properties of mean length of utterance in school-age children. *Language, Speech, and Hearing Services in Schools*, *53*(4), 1088–1100. https://doi.org/10.1044/2022_LSHSS-21-00115

Preston, J. L., & Seki, A. (2011). Identifying residual speech sound disorders in bilingual children: A Japanese-English case study. *American Journal of Speech-Language Pathology*, *20*(2), 73–85. https://doi.org/10.1044/1058-0360(2011/10-0057)

Pye, C., Mateo, P., Pfeiler, B., & Stengel, D. (2017). Analysis of variation in Mayan child phonologies. *Lingua*, *198*, 38–52. https://doi.org/10.1016/j.lingua.2017.07.001

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

RStudio Team. (2020). *RStudio: Integrated Development for R*. RStudio, PBC. http://www.rstudio.com

Rice, M. L., Smolik, F., Perpich, D., Thompson, T., Rytting, N., & Blossom, M. (2010). Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *Journal of Speech, Language, and Hearing Research*, *53*(2), 333–349. https://doi.org/10.1044/1092-4388(2009/08-0183)

Serratrice, L., Hesketh, A., & Ashworth, R. (2015). The use of reported speech in children's narratives: A priming study. *First Language*, *35*(1), 68–87. https://doi.org/10.1177/0142723715569552

Seung, H. M. (2022). *Smplot: Smplot—A package for statistical data visualization* [R package version 0.1.0]. https://www.smin95.com

Shipley, K. G., & McAfee, J. G. (2009). *Assessment in speech-language pathology: A resource manual* (4th ed). Delmar Cengage Learning.

Shriberg, L. D., & Kwiatkowski, J. (1982). Phonological disorders III: A procedure for assessing severity of involvement. *Journal of Speech and Hearing Disorders*, *47*, 256–270.

Shriberg, L. D., & Kwiatkowski, J. (1985). Continuous speech sampling for phonologic analyses of speech-delayed children. *Journal of Speech and Hearing Disorders*, *50*(4), 323–334. https://doi.org/10.1044/jshd.5004.323

Shriberg, L. D., Kwiatkowski, J., & Hoffmann, K. (1984). A procedure for phonetic transcription by consensus. *Journal of Speech, Language, and Hearing Research*, *27*(3), 456–465. https://doi.org/10.1044/jshr.2703.456

Shriberg, L. D., Tomblin, J. B., & McSweeny, J. L. (1999). Prevalence of speech delay in 6-year-old children and comorbidity with language impairment. *Journal of Speech, Language, and Hearing Research*, *42*(6), 1461–1481. https://doi.org/10.1044/jslhr.4206.1461

Skahan, S. M., Watson, M., & Lof, G. L. (2007). Speech-language pathologists' assessment practices for children with suspected speech sound disorders: Results of a national survey. *American Journal of Speech-Language Pathology*, *16*(3), 246–259. https://doi.org/10.1044/10580360(2007/029)

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders*, *55*(4), 779–798. https://doi.org/10.1044/jshd.5504.779

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *Journal of Neuroscience*, *32*(25), 8443–8453. https://doi.org/10.1523/JNEUROSCI.5069-11.2012

Soper, D. S. (2022). *Significance of the difference between two correlations calculator*. https://www.danielsoper.com/statcalc

Speech Accent Archive. (n.d.). Retrieved September 17, 2022, from https://accent.gmu.edu/

Stelmachowicz, P. G., Hoover, B. M., Lewis, D. E., Kortekaas, R. W. L., & Pittman, A. L. (2000). The relation between stimulus context, speech audibility, and perception for normal-hearing and hearing-impaired children. *Journal of Speech, Language, and Hearing Research*, *43*(4), 902–914. https://doi.org/10.1044/jslhr.4304.902

Stephen, L. (2017). Guatemalan immigration to Oregon: Indigenous transborder communities. *Oregon Historical Quarterly*, *118*(4), 554. https://doi.org/10.5403/oregonhistq.118.4.0554

Stipancic, K. L., Tjaden, K., & Wilding, G. (2016). Comparison of intelligibility measures for adults with Parkinson's disease, adults with multiple sclerosis, and healthy controls. *Journal of Speech, Language, and Hearing Research*, *59*(2), 230–238. https://doi.org/10.1044/2015_JSLHR-S-15-0271

Sturm, J. A., & Seery, C. H. (2007). Speech and articulatory rates of school-age children in conversation and narrative contexts. *Language, Speech, and Hearing Services in Schools*, *38*(1), 47–59. https://doi.org/10.1044/0161-1461(2007/005)

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *26*(2), 212–215. https://doi.org/10.1121/1.1907309

Sullivan, A. L. (2011). Disproportionality in special education identification and placement of English language learners. *Exceptional Children*, *77*(3), 317–334. https://doi.org/10.1177/001440291107700304

Thompson, S. (2021, May 27). The U.S. has the second largest population of Spanish speakers—How to equip your brand to serve them. *Forbes*. https://www.forbes.com/sites/soniathompson/2021/05/27/the-us-has-the-second-largest-population-of-spanish-speakers-how-to-equip-your-brand-to-serve-them/?sh=44a7d1a8793a

Tjaden, K., & Liss, J. M. (1995). The influence of familiarity on judgments of treated speech. *American Journal of Speech-Language Pathology*, *4*(1), 39–48. https://doi.org/10.1044/1058-0360.0401.39

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, *15*(4), 905–916. https://doi.org/10.1017/S1366728912000168

Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00577

Vaughn, C. R. (2019). Expectations about the source of a speaker's accent affect accent adaptation. *The Journal of the Acoustical Society of America*, *145*(5), 3218–3232. https://doi.org/10.1121/1.5108831

Weston, A. D., & Shriberg, L. D. (1992). Contextual and linguistic correlates of intelligibility in children with developmental phonological disorders. *Journal of Speech, Language, and Hearing Research*, *35*(6), 1316–1332. https://doi.org/10.1044/jshr.3506.1316

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag. http://ggplot2.org

Wickham, H., & Girlich, M. (2022). *tidyr: Tidy Messy Data* [R package version 1.2.0]. https://CRAN.R-project.org/package=tidyr

Wiig, E., Semel, Eleanor, & Secord, Wayne. (2013). *Clinical Evaluation of Language Fundamentals-Fifth Edition* (CELF-5). Pearson Assessments.

Wikipedia contributors. (2022, October). *Languages of the United States.* [Wikipedia]. https://en.wikipedia.org/wiki/Languages_of_the_United_States

Wolfe, V., Martin, D., Borton, T., & Youngblood, H. C. (2003). The effect of clinical experience on cue trading for the /r-w/ contrast. *American Journal of Speech-Language Pathology*, *12*(2), 221–228. https://doi.org/10.1044/1058-0360(2003/068)

Yeh, L.-L., & Liu, C.-C. (2021). Comparing the informativeness of single-word samples and connected speech samples in assessing speech sound disorders. *Journal of Speech, Language, and Hearing Research*, *64*(11), 4071–4084. https://doi.org/10.1044/2021_JSLHR-20-00172

Yoder, P. J., Woynaroski, T., & Camarata, S. (2016). Measuring speech comprehensibility in students with Down Syndrome. *Journal of Speech, Language, and Hearing Research*, *59*(3), 460–467. https://doi.org/10.1044/2015_JSLHR-S-15-0149

Zielinski, B. (2006). The intelligibility cocktail: An interaction between speaker and listener ingredients. *Prospect*, *3*, 20–43.